



INSTITUTE FOR DEFENSE ANALYSES

**Application of User-Oriented MOE to HPAC
Probabilistic Predictions of *Prairie*
Grass Field Trials**

Steve Warner, Project Leader
James F. Heagy
Nathan Platt

May 2001

Approved for public release;
distribution unlimited.

IDA Paper P-3586

Log: H 01-000105

Form SF298 Citation Data

Report Date <i>("DD MON YYYY")</i> 01052001	Report Type Final	Dates Covered (from... to) <i>("DD MON YYYY")</i> 01012000 01052001
Title and Subtitle Application of User-Oriented MOE to HPAC Probabilistic Predictions of Prairie Grass Field Trials		Contract or Grant Number
Authors Warner, Steve; Platt, Nathan; Heagy, James F.; et.al.		Program Element Number
Performing Organization Name(s) and Address(es) Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311-1772		Project Number
Sponsoring/Monitoring Agency Name(s) and Address(es) Dr. Allan Reiter Defense Threat Reduction Agency (DTRA) 6801 Telegraph Road Alexandria, VA 22310-3398		Task Number
Distribution/Availability Statement Approved for public release, distribution unlimited		Work Unit Number
Supplementary Notes		Performing Organization Number(s)
Abstract The goal of this task is to improve the verification, validation, and accreditation (VV&A) of hazard prediction and assessment models. These studies are part of a larger joint VV&A effort that DTRA and the Department of Energy, via the Lawrence Livermore National Laboratory (LLNL), are conducting. The Institute for Defense Analyses role includes conducting comparisons between models, providing analysis and discussions associated with these examinations, and exploring and developing measures of effectiveness (MOE) that can aid hazard prediction model validation and accreditation. This paper develops and demonstrates concepts that can ultimately aid validation and user accreditation of transport and dispersion models. The paper applies a user-oriented MOE to the probabilistic outputs that are obtained from the Hazard Prediction and Assessment Capability (HPAC) model predictions of the short-range Prairie Grass field trials. This MOE allows for the assessment of the probabilistic predictions in terms of the false negative and false positive fractions. Quantitative descriptions of a users potential risk tolerance are also described in this paper. This study develops these quantitative descriptions and uses them to evaluate the goodness of HPAC predictions of the Prairie Grass field trials from the perspective of a few notional users.		Monitoring Agency Acronym
		Monitoring Agency Report Number(s)

Subject Terms Model validation; hazardous material transport and dispersion; HPAC; probabilistic outputs; measure of effectiveness; Prairie Grass	
Document Classification unclassified	Classification of SF298 unclassified
Classification of Abstract unclassified	Limitation of Abstract unlimited
Number of Pages 304	

This work was conducted under contract DASW01 98 C 0067, Task DC-9-1797, for the Defense Threat Reduction Agency. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

© 2001 Institute for Defense Analyses, 1801 N. Beauregard Street, Alexandria, Virginia 22311-1772 • (703) 845-2000.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (NOV 95).

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-3586

**Application of User-Oriented MOE to HPAC
Probabilistic Predictions of *Prairie*
Grass Field Trials**

Steve Warner, Project Leader
James F. Heagy
Nathan Platt

PREFACE

The Institute for Defense Analyses (IDA) prepared this paper for the Defense Threat Reduction Agency (DTRA), in partial fulfillment of the task “Support for DTRA in the Validation Analysis of Hazardous Material Transport and Dispersion Prediction Models.” The objective of this effort was to conduct analyses and special studies associated with the verification, validation, and accreditation (VV&A) of hazardous transport and dispersion prediction models. This paper represents the third in a series that document our efforts associated with the above task. In particular, this paper compares probabilistic outputs of DTRA’s Hazard Prediction and Assessment Capability (HPAC) model to *Prairie Grass* field trial observations.

The IDA Technical Review Committee was chaired by Thomas P. Christie and consisted of J. Rosser Bobbitt, III, John Bombardt, Jr., Arthur Fries, Thomas R. King, and Douglas P. Schultz. The authors would like to thank Allan Reiter (DTRA), Leon Wittwer (DTRA), Scott Bradley (Logicon Advanced Technology), George Bieberbach (Logicon Advanced Technology), and Ian Sykes (ARAP Group, Titan Corporation) for their comments, critiques, and support throughout this effort.

APPLICATION OF USER-ORIENTED MOE TO HPAC PROBABILISTIC PREDICTIONS OF PRAIRIE GRASS FIELD TRIALS

TABLE OF CONTENTS

SUMMARY	1
A. Background	1
B. Purpose	2
C. User-Oriented MOE	2
D. Results And Discussion.....	3
1. The Two-Dimensional MOE Can Characterize HPAC Probabilistic Predictions	4
2. MOE Evaluation: User-Coloring of the Two-Dimensional Space.	8
E. Outline of This Paper	10
 1. INTRODUCTION	 1-1
A. User-Oriented Measure of Effectiveness	1-1
B. Brief HPAC Description	1-5
C. Description of the <i>Prairie Grass</i> Field Trial Experiment	1-7
References.....	1-R-1
 2. METHODOLOGIES.....	 2-1
A. Computation of 2D MOE.....	2-1
1. MOE Based on a Threshold.....	2-1
2. MOE Based on Summed Dosages	2-2
3. MOE Based on an Interpolated Area.....	2-3
a. Polar Bilinear.....	2-5
b. Delaunay Triangulation.....	2-6
4. Estimation of Confidence Intervals	2-6
B. Preparation of HPAC Probabilistic Output Predictions	2-7
C. <i>Prairie Grass</i> Data	2-10
1. Structure of Field Trial	2-10
2. Stability Category Groupings	2-12
3. Threshold for Sampler Observations	2-13
D. Quantitative Communication of Risk Tolerance: “User-Coloring”	2-14
1. Risk-Weighted Figure-of-Merit (R-FOM)	2-14
2. Area Size-Only FOM (A-FOM)	2-18
References.....	2-R-1

3.	RESULTS, ANALYSES, AND DISCUSSION.....	3-1
A.	Probabilistic Output Results: Probability of Exceeding a Threshold.....	3-1
1.	Across All Trials.....	3-1
2.	As a Function of Range	3-3
3.	As a Function of Stability Category	3-3
4.	Analogy to Receiver Operating Characteristic (ROC) Curve	3-6
B.	Probabilistic Output Results: Conditional Probability	3-8
1.	MOEs Based on Threshold of 60 mg-sec/m ³	3-8
2.	MOEs Based on Summed Dosage.....	3-13
C.	Probabilistic Output Results: Based on Interpolated Area.....	3-14
D.	MOE Values and Risk Tolerance: User-Colored Space	3-16
E.	Concept of Average Affected Area.....	3-19
	References.....	3-R-1

Appendix A – Acronyms

Appendix B – Some Comments on the Preparation of the HPAC
Predictions Used in This Analysis

Appendix C – HPAC Probabilistic Prediction Outputs Compared to
Prairie Grass Trials

Appendix D – HPAC Probabilistic Outputs: Full (0.50) Versus
Conditional (0.50) Probabilities for *Prairie Grass* Predictions

Appendix E – Results and Analyses: Supplemental Figures

Appendix F –Task Order Extract

LIST OF FIGURES

1.	Conceptual View of MOE Dimensions	2
2.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full Probability Predictions: All Trials.....	5
3.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full and Conditional Probability Predictions: All Trials.....	6
4.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full Probability Predictions: As a Function of Meteorological Stability Category Grouping	7
5.	Overlay of MOE Estimates (Based on 60 mg-sec/m ³ Threshold) for HPAC Full and Conditional Probability Predictions with All Trials: R-FOM (C _{FN} = 5, C _{FP} = 0.5) and A-FOM (s = 1.15).....	9
1-1.	Conceptual View of MOE Dimensions	1-2
1-2.	Key Characteristics of the Two-Dimensional MOE Space	1-3
1-3.	Interpretation of Comparisons: Exclusionary Zones	1-4
1-4.	Example Dosages (in mg-sec/m ³) and Field Trial Setup.....	1-8
2-1.	Illustration of Computations Based on a Threshold (Summed Distances Along the Arc)	2-2

2-2.	Illustration of Computations Based on Summed Dosages	2-2
2-3.	Example Dosages (in mg-sec/m ³) for <i>Prairie Grass</i> Field Trial 43.....	2-3
2-4.	Example Dosages (in mg-sec/m ³) for <i>Prairie Grass</i> Field Trial 43 Using Delaunay Triangulation Procedure to Perform Interpolation to a Regular Grid of Points.....	2-4
2-5.	Illustration of Polar Bilinear Interpolation	2-5
2-6.	Illustration of Conditional Probability Prediction Output: “Exceed. Value (V[P _c > 0.90])”.....	2-8
2-7.	Illustration of Full Probability Prediction Output: “Probability (P [V > 60 mg-sec/m ³])”	2-9
2-8.	<i>Prairie Grass</i> Trial 7 as a Function of Arc: Stability Category is 1	2-11
2-9.	Relationship Between R-FOM and 2D MOE: C _{FN} = 1, C _{FP} = 1.....	2-15
2-10.	Relationship Between R-FOM and 2D MOE: C _{FN} = 5, C _{FP} = 0.5.....	2-15
2-11.	User Coloring of MOE Space: R-FOM.....	2-17
2-12.	Illustration of Red-Green-Blue (RGB) Coloring Scheme	2-18
2-13.	A-FOM Isolines for Some Values of the Parameter <i>s</i>	2-19
2-14.	Graph of Function <i>f</i> Defined on [0, π/4].....	2-21
2-15.	Graph of A-FOM Function Defined on [0, π/2]	2-22
2-16.	Examples of A-FOM User-Coloring for <i>s</i> = 1.15, 1.5, and 2.....	2-22
3-1.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full Probability Predictions: All Trials.....	3-2
3-2.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full Probability Predictions: As a Function of Arc Range	3-4
3-3.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full Probability Predictions: As a Function of Stability Category Grouping (SCG)	3-5
3-4.	Notional ROC Curve: Trade-Off Between Detection and False Alarm Probabilities for a Given Sensor and Stimulus with Various Point on the Blue Line Corresponding to Different Decision Behavior	3-7
3-5.	MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full and Conditional Probability Predictions: All Trials	3-8
3-6.	Direct Comparison of HPAC Full (Red Regions) and Conditional Probability (Blue Regions) Predictions (with Mean Value Prediction in Yellow): MOE Estimates (Based on 60 mg-sec/m ³) for All Trials	3-9
3-7.	0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Full Probability of Exceeding Threshold Value for Trial 58: Stability Class is 6.....	3-10
3-8.	0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Full Probability of Exceeding Threshold Value for Trial 16: Stability Class is 1.....	3-11
3-9.	Direct Comparison of HPAC Full (Red Regions) and Conditional Probability (Blue Regions) Predictions (with Mean Value Prediction in Yellow): MOE Estimates (Based on 60 mg-sec/m ³) for the 50- and 800- Meter Arc.....	3-12
3-10.	MOE Estimates (Based on Summed Dosage) for HPAC Conditional Probability Predictions: All Trials	3-13

3-11.	Comparison of “Summed Distance Along the Arc” and Interpolation Techniques: MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Full Probability Predictions and All Trials	3-14
3-12.	Comparison of “Summed Distance Along the Arc” and Interpolation Techniques: MOE Estimates (Based on 60 mg-sec/m ³) for HPAC Conditional Probability Predictions and All Trials	3-15
3-13.	Comparison of “Summed Dosage Along the Arc” and Interpolation Techniques: MOE Estimates (Based on Summed Dosage – Linear Scale) for HPAC Conditional Probability Predictions and All Trials	3-15
3-14.	Overlay of MOE Estimates (Based on 60 mg-sec/m ³ Threshold) for HPAC Full Probability Predictions and All Trials: R-FOM	3-16
3-15.	Overlay of MOE Estimates (Based on 60 mg-sec/m ³ Threshold) for HPAC Full Probability Predictions and All Trials: A-FOM	3-17
3-16.	Overlay of MOE Estimates (Based on 60 mg-sec/m ³ Threshold) for HPAC Full and Conditional Probability Predictions with All Trials: R-FOM (C _{FN} = 5, C _{FP} = 0.5) and A-FOM (s = 1.15)	3-18
3-17.	Overlay of MOE Estimates (Based on 60 mg-sec/m ³ Threshold and Polar Bilinear Interpolation) for HPAC Full Probability Predictions and All Trials: R-FOM (C _{FN} = 2, C _{FP} = 0.5)	3-19
3-18.	Two Notional Scenarios with “Identical” MOE Values	3-20
3-19.	Average Affected Areas as a Function of Stability Category and Based on a 60 mg-sec/m ³ Threshold and Delaunay Triangulation Interpolation	3-21
3-20.	3-Dimensional Locations (MOE and Average Affected Area) as a Function of Stability Category and Based on a 60 mg-sec/m ³ Threshold and Delaunay Triangulation Interpolation for 49 Individual Prairie Grass Field Trial Predictions	3-23
3-21.	3-Dimensional Locations (MOE and Average Affected Area) as a Function of Stability Category and Based on a 60 mg-sec/m ³ Threshold and Delaunay Triangulation Interpolation for 49 Individual Prairie Grass Field Trial Predictions: “End-On” Views	3-24
3-22.	MOE (95 th Percent Confidence) Regions and Average Affected Area as a Function of Stability Category and Based on a 60 mg-sec/m ³ Threshold and Delaunay Triangulation Interpolation for the Prairie Grass Field Trial Predictions	3-26

LIST OF TABLES

1-1.	List of <i>Prairie Grass</i> Trials That Were Used	1-9
2-1.	Distance Between Samplers by Arc	2-11
2-2.	Stability Category Groupings (SCGs) Used in This Study	2-12
2-3.	Sample Sizes Used for Model Comparisons to <i>Prairie Grass</i> Field Trials...	2-13

SUMMARY

A. BACKGROUND

During fiscal year 2000, a series of studies in support of the Defense Threat Reduction Agency (DTRA) was begun. The goal of these studies is to improve the verification, validation, and accreditation (VV&A) of hazard prediction and assessment models and capabilities. These studies are part of a larger joint VV&A effort that DTRA and the Department of Energy, via the Lawrence Livermore National Laboratory (LLNL), are conducting. This joint effort includes comparisons of the LLNL and DTRA transport and dispersion (T&D) modeling systems, NARAC and HPAC, respectively.¹

IDA's role is to conduct independent analysis and special studies associated with this VV&A effort. This role includes conducting comparisons between the models, providing analysis and discussions associated with these examinations, and exploring and developing measures of effectiveness (MOEs) that can aid hazard prediction model validation and accreditation.²

This is the third paper in a series that documents our efforts associated with this task. The first paper in this series developed novel user-oriented MOEs and applied them to the comparison of HPAC and NARAC predictions of the 1956 *Prairie Grass* field trial.³ The focus of the second paper was on model-to-model comparisons for a collection of relatively simple, controlled, release scenarios.⁴

¹ NARAC = National Atmospheric Release Advisory Center and HPAC = Hazard Prediction and Assessment Capability.

² Appendix F of this document contains an extract from the pertinent Fiscal Year 2000 task order.

³ Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T., and Larson, D., User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models, IDA Paper P-3554, January 2001.

⁴ Warner S., Heagy, J. F., Platt, N., Larson, D., Sugiyama, G., Nasstrom, J. S., Foster, K. T., Bradley, S., and Bieberbach, G., Evaluation of Transport and Dispersion Models: An Initial Comparison of Hazard Prediction and Assessment Capability (HPAC) and National Atmospheric Release Advisory Center (NARAC) Predictions, IDA Paper P-3555, May 2001.

B. PURPOSE

The purpose of this paper is to develop and demonstrate concepts that can *ultimately* aid validation and user accreditation of transport and dispersion models. This paper applies the user-oriented MOE to the probabilistic outputs that are obtained from HPAC predictions of the short-range *Prairie Grass* field trials. Thus, this study extends the application of the user-oriented MOE. In addition, this paper considers the quantitative communication of risk tolerance with respect to assessing the magnitudes of (that is, evaluating) particular MOE estimates.

C. USER-ORIENTED MOE

A fundamental feature of any comparison of hazard prediction model output to observations is the over- and underprediction regions. We define the *false negative* region where a hazard is observed but not predicted, and the *false positive* region where a hazard is predicted but not observed. Figure 1 shows the observed and predicted area above some nominal dosage level threshold. Numerical estimates of the false negative region (A_{FN}), the false positive region (A_{FP}), and the overlap region (A_{OV}) characterize this conceptual view.

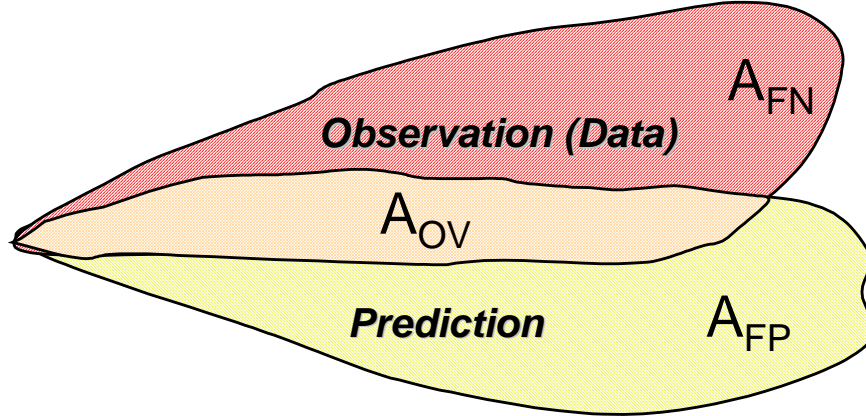


Figure 1. Conceptual View of MOE Dimensions

The MOE that we consider for this study has two dimensions. The x-axis corresponds to the ratio of overlap area to observed area and the y-axis corresponds to the ratio of overlap area to predicted area. When these mathematical definitions are algebraically rearranged (Equation 1 below), we recognize that the x-axis corresponds to *1 minus the false negative fraction* and the y-axis corresponds to *1 minus the false positive fraction*.

$$MOE = \left(\frac{A_{OV}}{A_{OB}}, \frac{A_{OV}}{A_{PR}} \right) = \left(\frac{A_{OV}}{A_{OV} + A_{FN}}, \frac{A_{OV}}{A_{OV} + A_{FP}} \right) = \left(1 - \frac{A_{FN}}{A_{OB}}, 1 - \frac{A_{FP}}{A_{PR}} \right) \quad (1)$$

where A_{FN} = area of false negative, A_{FP} = area of false positive, A_{OV} = area of overlap, A_{PR} = area of the prediction, and A_{OB} = area of the observation.

MOE estimates based on HPAC probabilistic predictions of the *Prairie Grass* field trials were used in this study. *Prairie Grass* field trials were conducted during the summer of 1956 in north central Nebraska. The primary objective of *Project Prairie Grass* was to determine the rate of diffusion of a neutrally buoyant tracer gas as a function of meteorological conditions. These experiments involved continuous 10-minute releases of sulfur dioxide (SO_2) from a near-surface point source. Downwind SO_2 concentrations were sampled along five concentric, semi-circular arcs located 50, 100, 200, 400, and 800 meters away from the gas source.

D. RESULTS AND DISCUSSION

The major results of this study can be characterized as follows.

- This paper demonstrates that the two-dimensional MOE can be used to describe the performance of HPAC probabilistic predictions. The MOE allows for the assessment of the probabilistic predictions in terms of the false negative and false positive fractions. HPAC mean value predictions can be interpreted in an identical manner and thus, importantly, can be assessed side-by-side with probabilistic prediction outputs. The computed MOE estimates support the notion that, by adjusting the input probability values, in general, HPAC can generate a wide range of associated false positive and false negative fractions.
- Quantitative descriptions of a user's potential risk tolerance are elucidated in this paper. This study develops these quantitative descriptions and uses them to evaluate the "goodness" of HPAC predictions of the *Prairie Grass* field trials from the perspective of a few notional users. Given a quantitative description of a user's risk tolerance, one then can select HPAC probability values that are acceptable.

The rest of this section provides evidence for the above assertions.

1. The Two-Dimensional MOE Can Characterize HPAC Probabilistic Predictions

Figure 2 shows approximate 95th percent confidence regions for MOE estimates⁵ based on HPAC full probability⁶ and mean value (the yellow region) predictions of the 51 *Prairie Grass* field trials. If we define a hazard area as wherever the dosage exceeds some critical threshold (e.g., 60 mg-sec/m³),⁷ then we can interpret the 0.01 full probability prediction as encompassing the area that has a 0.99 (1.00-0.01) probability of containing the hazard area. The 0.01 full probability values are always associated with the widest predictions (most conservative) and the 0.999 full probability values are always associated with the narrowest predictions.

The MOE indicates that, in going from probability values of 0.999 to 0.01, the false negative fraction decreases and the false positive fraction increases. With the exception of the estimate based on the 0.01 probability prediction, the chosen probabilistic values (0.50, 0.80, 0.85, 0.90, 0.95, and 0.999) led to increased false negative and decreased false positive fractions, relative to the mean value prediction.

Figure 3 compares HPAC “full” and “conditional” probabilistic predictions using the MOE. The relatively conservative nature of the conditional predictions is illustrated by their associated decreased false negative *and* increased false positive fractions for the corresponding probabilities. The biggest differences between full and conditional probability MOE estimates are seen at the 50-meter arc. For the 800-meter arc, the conditional probability predictions are much more similar to the full probability results.⁸

⁵ The MOE estimates discussed in this section are based on a threshold of 60 mg-sec/m³. The approximate 95th percent confidence regions were computed using a bootstrap (resampling) technique. These regions quantify the uncertainty associated with the MOE point estimate. See Chapter 2, Section A, for more information.

⁶ For this study, we examined “full” and “conditional” probabilities. Basically, HPAC full probability predictions include all plume “realizations,” whereas conditional probabilistic predictions exclude intermittent spatial distributions, that is, the predicted zero dosage values at some locations. In this way, the HPAC conditional probabilistic output is meant to provide a somewhat conservative prediction. See Chapter 2, Section B, for additional details.

⁷ Chapter 2, Section C.3 discusses sampler threshold limits and the choice of 60 mg-sec/m³.

⁸ See Figure 3-9 for comparisons of full and conditional probability results at 50- and 800-meters.

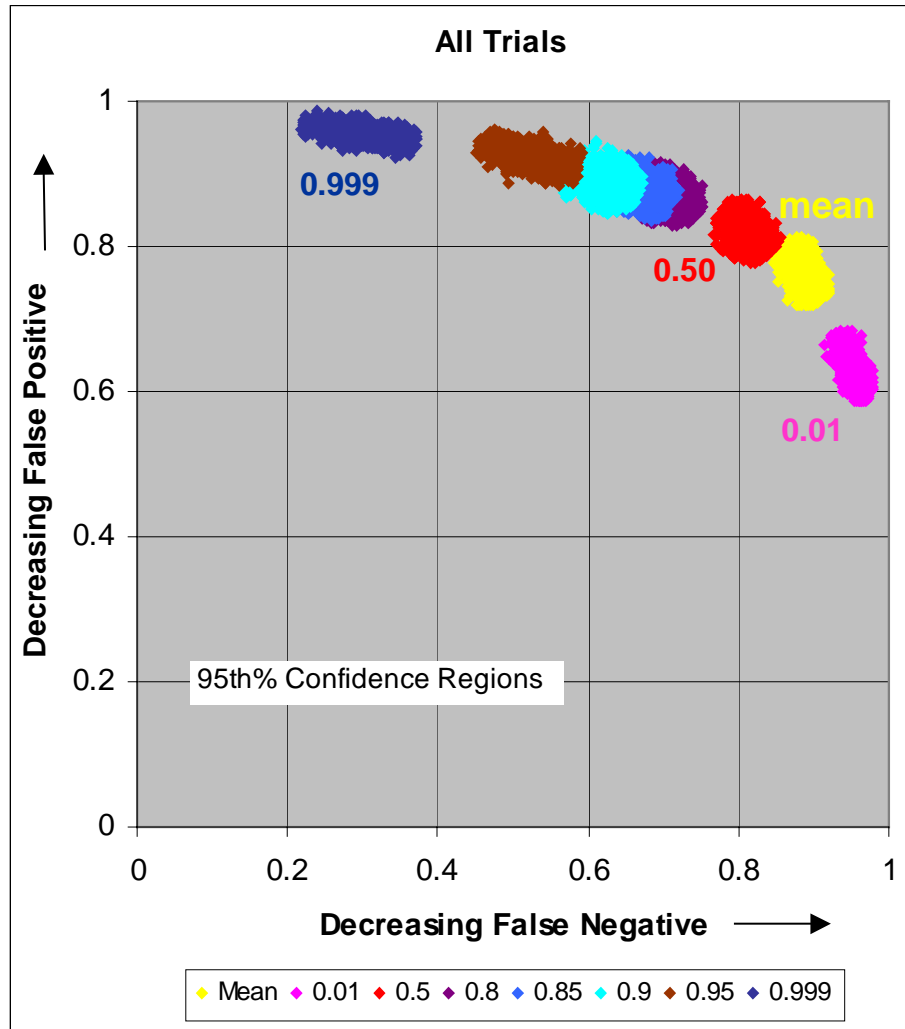


Figure 2. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full Probability Predictions: All Trials

The exclusion of intermittent spatial distributions, that is, the predicted zero dosage values at some locations, increases the conditional probabilities relative to the full probabilities. For these data, the increases are maximized at shorter ranges and, for the investigated threshold value, lead to wider predicted plumes (relative to the full probability predictions).

The “best” HPAC predictions of the 51 *Prairie Grass* field trials, that is, those that led to the MOE value closest to the perfect (1,1) point, correspond to the 0.50 full probability predictions. The 0.50 full probability-based MOE is just barely closer to (1,1) than the MOE estimate associated with the mean value predictions.

Figures 2 and 3 show similar trends in MOE value as a function of the selected probability level. As the probability value is lowered, from 0.999 to 0.01, the plume at

60 mg-sec/m³ gets larger, and, hence, the false negative fraction decreases and the false positive fraction increases. The two-dimensional MOE used in this study provides a direct way to communicate the trade-off associated with different choices for the probability values.⁹

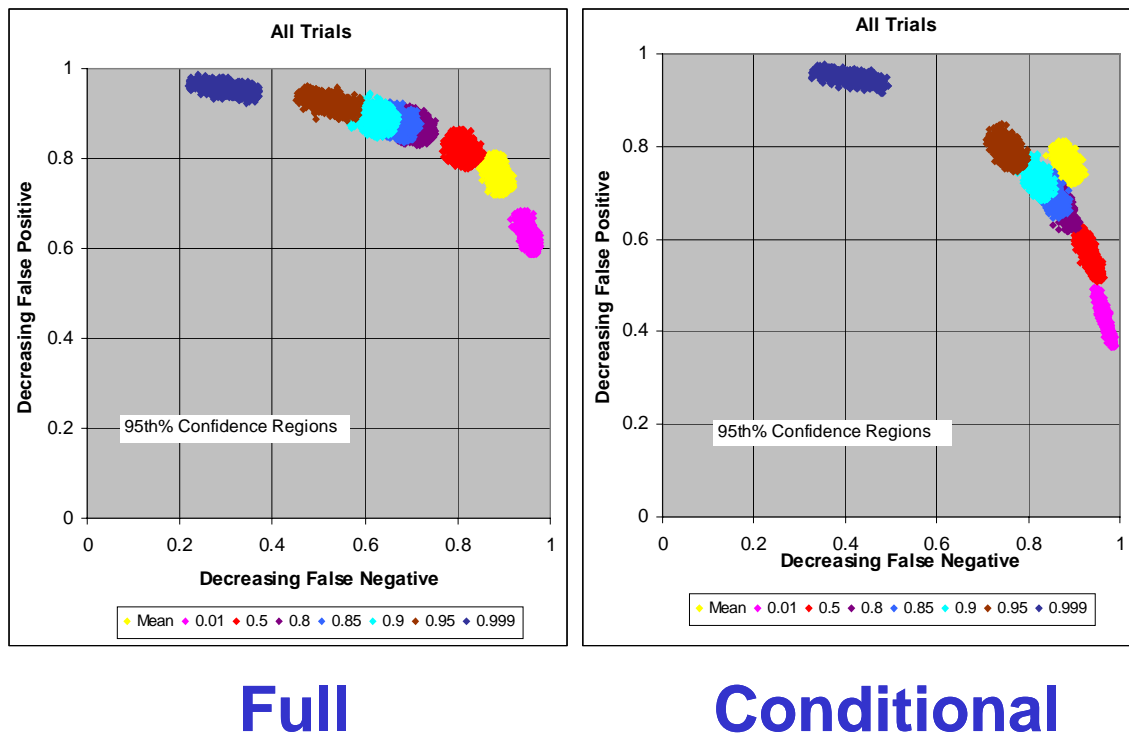
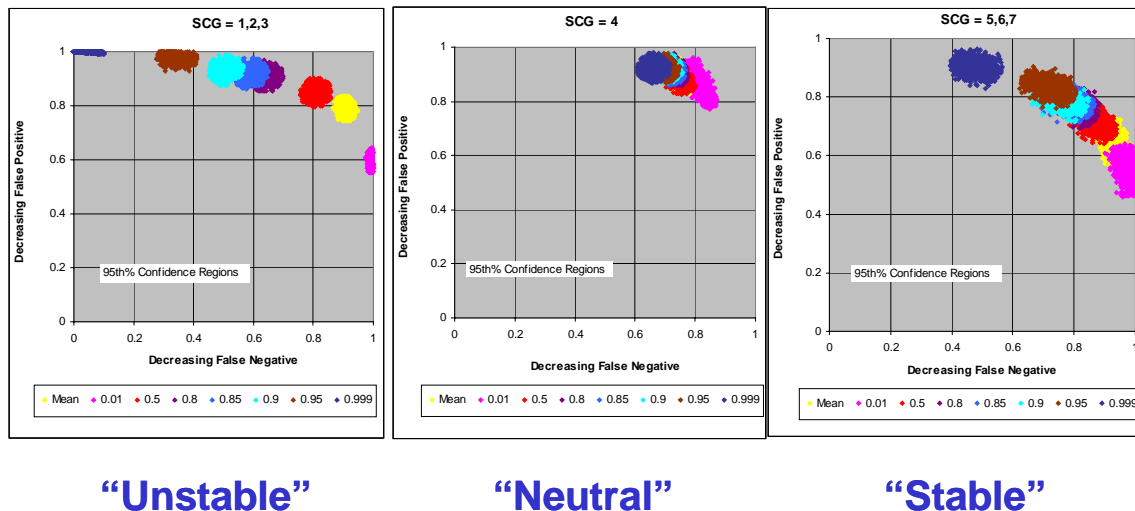


Figure 3. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full and Conditional Probability Predictions: All Trials

We also examined MOE values for HPAC probabilistic predictions as a function of arc range and meteorological stability category. In general, we found the same trends as described above. Figure 4 points out an important feature associated with the HPAC predictions of the “neutral” *Prairie Grass* field trials.¹⁰

⁹ In Chapter 3, Section A.4, we draw an analogy to this behavior with a classic detection theory paradigm that defines the Receiver Operating Characteristic curve as a function that relates the probability of detection to the probability of false alarm.

¹⁰ We used stability category assignments, 1 through 7, based on a recent study. See Irwin, J.S., and Rosu, M-R., “Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models,” *Proceedings of the 10th Joint Conference on the Applications of Air Pollution Meteorology*, American Meteorological Society, Boston, pp. 6-10, 1998. The SCG of Figure 4 are as follows: unstable = SCG 1,2,3; neutral = SCG 4; and stable = SCG 5,6,7.



**Figure 4. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full Probability Predictions:
As a Function of Meteorological Stability Category Grouping**

For the unstable and stable trials, MOE values traverse a wide range in two-dimensional space as a function of the chosen probability. For example, for the unstable trials, the false negative fraction varies from about 1.0 to 0.0. Similarly, for the stable trials, the false positive fraction varies from about 0.50 to 0.90. However, for the neutral trials, there is very little “movement” in the MOE space. For the neutral trials, the false negative fraction varies from about 0.65 to 0.85 and the false positive fraction varies from about 0.80 to 0.90. The relative lack of variability as a function of chosen probability may be a manifestation of the physical properties associated with neutral stability conditions. The suggestion is, that for these predictions, a user would not have much choice with respect to false negative/positive fractions for the neutral conditions. It is plausible that this lack of variability as a function of selected probability value for the neutral trials is also related to the adopted protocol used for these short-range comparative predictions.

We also computed MOE values after applying area interpolation routines to the observed data (and predictions in the case of the full probabilities).¹¹ We found that, for these short-range, densely sampled observations, the results based on interpolation were almost identical to the baseline results (reported above). The value of the interpolation routine is that it allows us to compute MOE values based on actual areas (e.g., in km²). MOE values based on actual areas have the advantage of direct scaling with an assumed

¹¹ We examined two interpolation techniques, Delaunay triangulation and polar bilinear, in this study. See Chapter 2, Section A.3, and Chapter 3, Section C, for additional details.

underlying spatially uniform population, therefore allowing for the interpretation of the MOE in terms of an affected population.

2. MOE Evaluation: User-Coloring of the Two-Dimensional Space

Typical operational users of transport and dispersion models might consider false positives and false negatives quite differently. For many applications, false positives would be much more acceptable to the user than false negatives (which could result in decisions that directly lead to death or injury). Equation 2 describes an example of a possible user scoring function. Basically, Equation 2 describes a spatial figure-of-merit (FOM) that evaluates the ratio of overlap area to total (predicted plus observed) area, for example, above a specified threshold.¹² The coefficients, C_{FN} and C_{FP} , weight the false negative and false positive regions, respectively, according to a given user's acceptable risk tolerance. Hence, we refer to this notional user scoring function as the Risk-Weighted FOM (R-FOM).

$$R-FOM = \frac{A_{OV}}{(A_{OV} + C_{FN}A_{FN} + C_{FP}A_{FP})} \quad (2)$$

where $C_{FN}, C_{FP} > 0$.

A transport and prediction model might be applied in such a way that the actual location of the hazard or direction of the plume was of no particular importance. For example, such a model might be used to study potential future outcomes of an accidental or intentional release. In this case, the actual weather (e.g., wind speed and direction) of the future (for instance, one year in the future) cannot be known with any accuracy. This researcher may use a scoring function that simply compares the sizes of the predicted and observed areas.

For the above notional user, we also developed an area size-only scoring function, denoted A-FOM, by considering the diagonal in the two-dimensional MOE space. This diagonal consists of the points that incorporate "equal size" predictions and observations. First, let us assume a hypothetical requirement that A_{PR} and A_{OB} must be within a factor of s of each other, with $s > 1$. Then, this can be stated mathematically by requiring that $1/s \leq A_{PR}/A_{OB} \leq s$. This corresponds to our A-FOM scoring function.

¹² A related spatial FOM that does not include explicit weighting coefficients was described in a recent study. See Mosca, S., Graziani, G., Klug, W., Bellasio, R., and Bianconi, R., "A Statistical Methodology for the Evaluation of Long-Range Dispersion Models: An Application to the ETEx Exercise," *Atmos. Environ.*, Vol. 32, No. 24, 4307-4324, 1998.

The above R-FOM and A-FOM scoring functions can be mapped onto the two dimensional MOE space via red-green-blue coloring.¹³ This mapping provides a convenient way to communicate an assessment of a specific MOE value (i.e., green is acceptable and red is unacceptable). Figure 5 overlays MOE confidence regions for HPAC predictions of the *Prairie Grass* field trials on notional user-coloring schemes based on R-FOM and A-FOM.

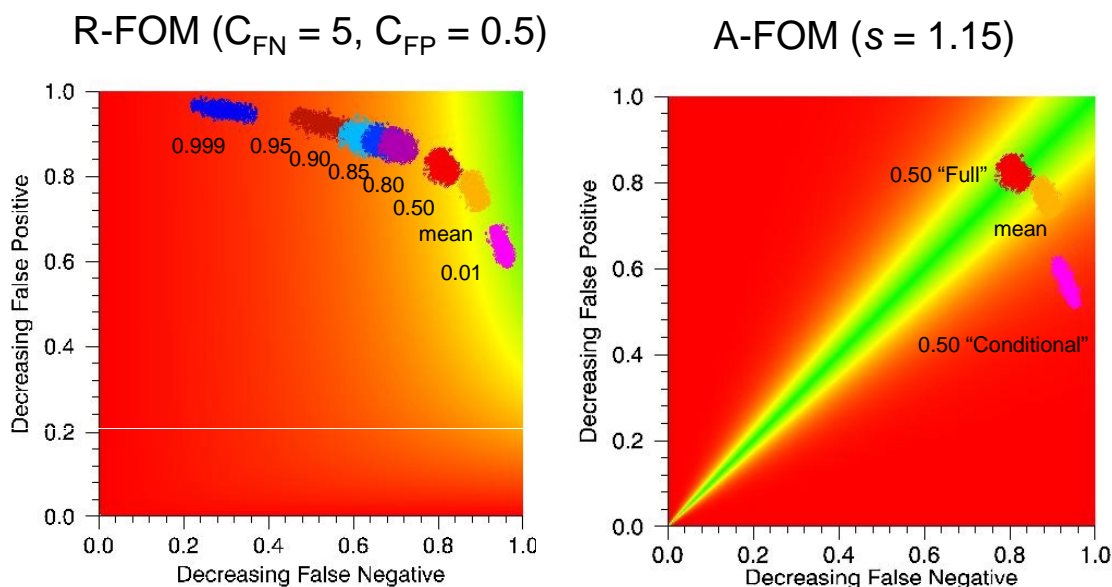


Figure 5. Overlay of MOE Estimates (Based on 60 mg-sec/m³ Threshold) for HPAC Full and Conditional Probability Predictions with All Trials:
R-FOM ($C_{FN} = 5$, $C_{FP} = 0.5$) and A-FOM ($s = 1.15$)

The following evaluations result from the notional user-coloring/scoring schemes illustrated in figure 5. For the conservative R-FOM requirement ($C_{FN} = 5$, $C_{FP} = 0.5$), the 0.01 full probability predictions perform best. For the A-FOM requirement that the predicted and observed hazard area sizes at 60 mg-sec/m³ be within 15 percent of each other, the 0.50 full probability prediction outperforms the mean value and 0.50 conditional probability results.¹⁴

¹³ See Chapter 2, Section C, for additional details.

¹⁴ In this paper, we also introduce the concept of averaged affected area (Chapter 3, Section E). Basically, we define the averaged affected area as one-half of the sum of the observed and predicted areas. This concept can be used to develop the absolute value (in terms of “hazard” area (km²) or number of affected people) associated with a given MOE estimate.

The above notional user scoring functions, R-FOM and A-FOM, represent illustrative examples. We expect that the evaluation of a model’s predictive capability, via the MOE, would involve a particular user in the formulation of his/her specific requirement (i.e., “user coloring”).

E. OUTLINE OF THIS PAPER

This paper is divided into three chapters and six appendices. Chapter 1 provides a brief introduction and describes the user-oriented MOE that we use in this study. Chapter 2 describes this study’s methodologies. In particular, the computation of the MOE, a few details of the *Prairie Grass* field trials, and HPAC probabilistic prediction outputs are described in Chapter 2. Chapter 3 evaluates HPAC probabilistic predictions of the *Prairie Grass* field trials in terms of the MOE and provides analyses and discussion.

Appendix A provides an acronym list. Appendix B documents some minor issues (including software “bugs”) that arose during the preparation of HPAC probabilistic prediction outputs. Appendices C and D provide comparisons, by sampler arc, of *Prairie Grass* field trial observations and model predictions. Supplemental figures, which show MOE confidence regions for HPAC predictions of the *Prairie Grass* field trials, are deposited in Appendix E. Finally, Appendix F includes an extract of the task order associated with this study.

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

A series of studies in support of the Defense Threat Reduction Agency (DTRA) was begun in fiscal year 2000. The goal of these studies is to improve the verification, validation, and accreditation (VV&A) of hazard prediction and assessment models and their capabilities.¹ These studies are part of a larger joint VV&A effort that DTRA and the Department of Energy (DOE), via the Lawrence Livermore National Laboratory (LLNL), are conducting. This joint effort includes comparisons of the LLNL and DTRA transport and dispersion (T&D) modeling systems, NARAC and HPAC, respectively, and their predictions. IDA's role is to conduct independent analysis and special studies associated with this VV&A effort. This role includes conducting comparisons between the models, providing analysis and discussions associated with these examinations, and exploring and developing measures of effectiveness (MOEs) that can aid hazard prediction model validation and accreditation.²

This is third paper in a series that documents our efforts associated with this task. The first paper in this series developed novel user-oriented MOEs and applied them to the comparison of HPAC and NARAC predictions of the 1956 *Prairie Grass* field trial [Ref. 1-1]. The focus of the second paper was on model-to-model comparisons for a collection of relatively simple, controlled, release scenarios [Ref. 1-2]. This paper applies the user-oriented MOEs to the probabilistic outputs that are obtained from HPAC predictions of the 1956 *Prairie Grass* field trial. In addition, this paper considers the quantitative communication of risk tolerance with respect to assessing the magnitudes of particular MOE estimates.

A. USER-ORIENTED MEASURE OF EFFECTIVENESS

In 1999, IDA conducted a study for the Office of the Secretary of Defense that, in part, explored military user requirements for hazard predictions [Ref. I-3]. During that

¹ NARAC = National Atmospheric Release Advisory Center and HPAC = Hazard Prediction and Assessment Capability.

² Appendix F of this document contains an extract from the pertinent Fiscal Year 2000 task order.

study, it was apparent that there was a need for measures that clearly communicate to the user the relative worth of a model's predictions.

Previously, we developed and described a user-oriented MOE [Ref. 1-1]. A fundamental feature of any comparison of hazard prediction model output to observations is the over- and underprediction regions. We define the *false negative* region where a hazard is observed but not predicted, and the *false positive* region where a hazard is predicted but not observed. Figure 1-1 shows the observed and predicted area at the same dosage level for some nominal situation. Numerical estimates of the false negative region (A_{FN}), the false positive region (A_{FP}), and the overlap region (A_{OV}) characterize this conceptual view.

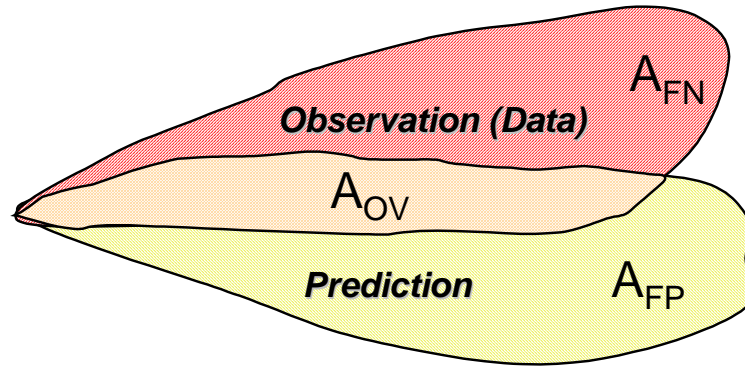


Figure 1-1. Conceptual View of MOE Dimensions

The MOE that we considered has two dimensions. The x-axis corresponds to the ratio of overlap area to observed area and the y-axis corresponds to the ratio of overlap area to predicted area. When these mathematical definitions are algebraically rearranged (Equation 1-1 below), we recognize that the x-axis corresponds to *1 minus the false negative fraction* and the y-axis corresponds to *1 minus the false positive fraction*.

$$MOE = \left(\frac{A_{OV}}{A_{OB}}, \frac{A_{OV}}{A_{PR}} \right) = \left(\frac{A_{OV}}{A_{OV} + A_{FN}}, \frac{A_{OV}}{A_{OV} + A_{FP}} \right) = \left(1 - \frac{A_{FN}}{A_{OB}}, 1 - \frac{A_{FP}}{A_{PR}} \right) \quad (1-1)$$

where A_{FN} = area of false negative, A_{FP} = area of false positive, A_{OV} = area of overlap, A_{PR} = area of the prediction, and A_{OB} = area of the observation. Consistent with the above algebraic rearrangement, Figure 1-2 shows the area of false negative decreasing from left to right.

Figure 1-2 demonstrates some of the key characteristics of the two-dimensional (2D) MOE space. We begin with the (1,1) point located at the upper-right corner. Here, both plumes overlap entirely (no false negative nor false positive fraction), and thus the

model would achieve perfect agreement with the field trial. Point (0,0) signifies that there is no region of overlap, and thus the model disagrees completely with the field trial. This 2D MOE includes directional effects; that is, the prediction of the location of a hazard and not just the shape and size of the plume is critical to obtaining a high MOE “score.”

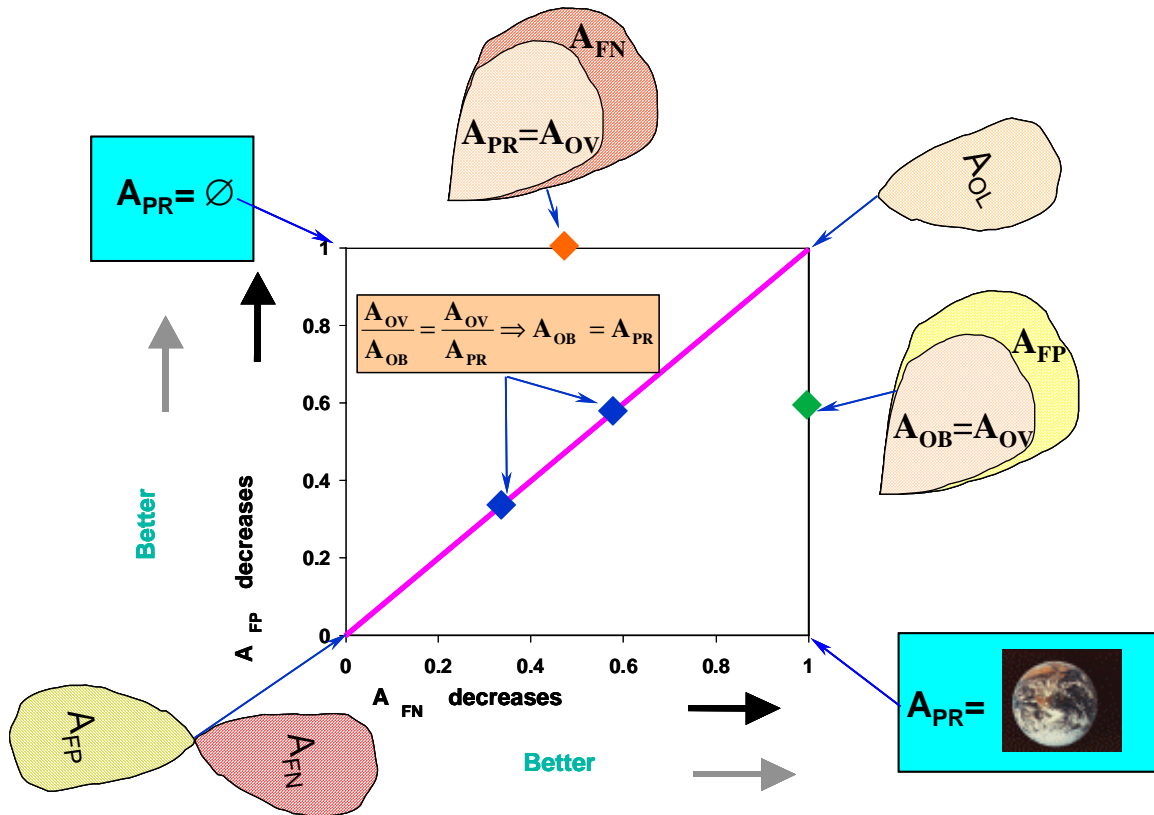


Figure 1-2. Key Characteristics of the Two-Dimensional MOE Space

Point (1,0) represents a situation where there is no false negative region, while there is an “infinite” false positive region (for nonzero releases). At this point, the implication is that the model predicts hazard everywhere. Along the line, $x = 1$, the prediction completely envelops the observation.

Point (0,1) signifies that there is no false positive region, but there is an “infinite” false negative region (for nonzero releases). At this point, the implication is that the model predicts no hazard. Along the line, $y = 1$, the observation completely envelops the prediction.

The “purple” diagonal line represents the situation where the prediction and the observation have identical “total” areas. As one traverses this diagonal line from (1,1)

toward (0,0), the fraction of overlap area between the predicted and observed plumes decreases.

Figure 1-3 suggests an additional interpretation of the 2D MOE. In this figure, the gold region represents the estimate of the MOE for some set of fictional model predictions and field trial observations. The point estimate, perhaps the vector mean value of several similar trials, would be found at the center of this region, and the overall size of the region represents the uncertainty associated with the point estimate of the MOE.

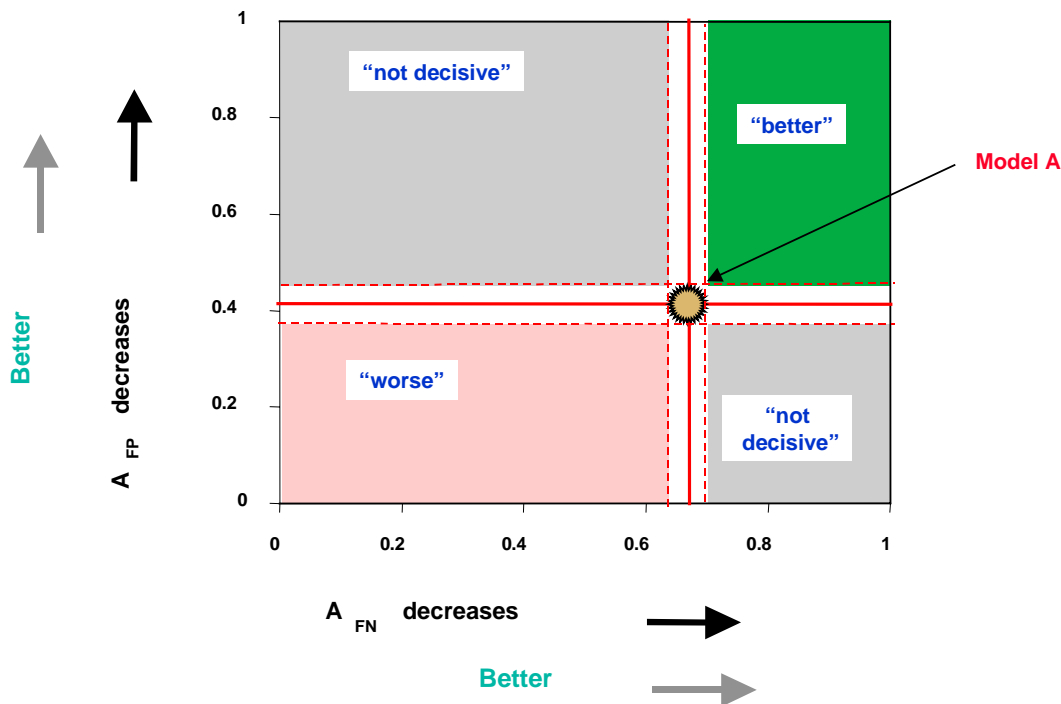


Figure 1-3. Interpretation of Comparisons: Exclusionary Zones

If a second set of model predictions were compared to “Model A,” several conclusions might be anticipated. The second model’s MOE estimate might be found in the region shaded “pink” (lower left). This would imply that Model A performs significantly better; both its false positive and false negative fractions are lower. Alternatively, the second model might lead to an estimate in the green region (upper right) – an indication that Model A is the poorer performer (for this set of field trial observations).

Finally, the new model predictions might lead to an MOE value that is located within the gray region. The implication here is that a user would have to make a

determination as to the tradeoff between false positive and false negative before deciding which model was most appropriate for his or her specific application.

B. BRIEF HPAC DESCRIPTION

HPAC is composed of a suite of software modules that can generate source terms for hazardous releases, retrieve and prepare meteorological information for use in a prediction, model the T&D of the hazardous release over time, and plot and report the results of these calculations. HPAC has been applied to various national defense problems including military studies and operational planning.³

For hazardous material T&D, HPAC uses the Second-Order Closure Integrated Puff (SCIPUFF) model and an associated mean wind field model. SCIPUFF, which is a Lagrangian model for atmospheric dispersion that uses the Gaussian puff numerical method – an arbitrary time-dependent concentration field is represented by three-dimensional Gaussian distributions – bases its turbulent diffusion parameterization on second-order closure theories. This methodology provides a link between measurable velocity statistics and the predicted dispersion rates. This “second-order” feature implies that concentration variance can also be computed, and this uncertainty estimate can be used as the basis for a probabilistic description of the dispersion prediction.⁴

HPAC is capable of generating three general types of dosage output: mean predictions, “full” probabilistic predictions, and “conditional” probabilistic predictions. These three prediction types are discussed below.

At each point in space and time SCIPUFF, the plume propagation methodology within HPAC, furnishes estimates of the mean concentration, μ , and the standard deviation in the concentration, σ . A “mean” prediction in HPAC is simply a plot of the mean concentration contours (or dosage contours) at a given snapshot in time.

Uncertainties in weather conditions (due primarily to imperfect wind field descriptions drawn from a limited number of observations), coupled with the random nature of the turbulent diffusion process, lead to uncertainties in the estimates of μ and σ . HPAC attempts to account for these uncertainties by using the empirically-based “clipped

³ This study examines the most recently released HPAC software version, 3.2.1 [Ref. 1-4].

⁴ See Reference 1-5 for several reports that provide details of HPAC design, functionality, capabilities, and V&V.

normal” probability density function (Equation 1-2) to characterize concentration fluctuations [Ref. 1-6],

$$\rho(c) = \gamma\delta(c) + \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right), \quad (1-2)$$

where μ and σ are the mean and standard deviation estimates. The term $\delta(c)$ is a *delta function*, which indicates there is some probability of finding a *zero* concentration value at the specified point in space and time. The parameter γ is known as the *intermittency*, and gives the probability of obtaining a zero concentration value (Equation 1-3); γ is related to μ and σ through

$$\gamma = \frac{1}{2} \left(1 - \operatorname{erf}\left(\frac{\mu}{\sigma\sqrt{2}}\right)\right). \quad (1-3)$$

Note that $0 < \gamma < 1/2$, which indicates there is always some probability of a non-zero concentration value. Armed with this empirical characterization, one can then ask the two following questions:

- (1) What is the probability that the concentration exceeds a given threshold value? In the language of the HPAC system, this is known as a “full” probability calculation. For a given snapshot in time, the full probability calculation yields *probability contours* in space that satisfy the threshold criterion. Contours associated with smaller probability values generally occupy larger regions.
- (2) Given that the concentration is non-zero, what is the concentration, such that the probability of exceeding this value is a predetermined value p , where $0 < p < 1$. This is known as a “conditional” probability calculation (one is asking for a concentration, conditioned on the requirement that it is non-zero). At a given snapshot in time, the conditional probability calculation yields *concentration contours* in space, all of which satisfy the probability requirement. In this case the contour associated with a given concentration value, will generally occupy a larger region as p decreases.

One can ask the same two questions for dosages (i.e., integrated concentrations), which are the focus of the comparisons in this study. In particular, this study uses the previously described 2D MOE to compare these HPAC probabilistic outputs for predictions of the *Prairie Grass* field trials.

C. DESCRIPTION OF THE *PRAIRIE GRASS* FIELD TRIAL EXPERIMENT

Prairie Grass field trials were conducted during the summer of 1956 in north central Nebraska near the town of O'Neil.⁵ The primary objective of *Project Prairie Grass* was to determine the rate of diffusion of a neutrally buoyant tracer gas as a function of meteorological conditions. These experiments involved continuous 10-minute releases of sulfur dioxide (SO₂) from a near-surface point source. Downwind SO₂ concentrations were sampled along five concentric, semi-circular arcs located 50, 100, 200, 400, and 800 meters away from the gas source.

The sampling network utilized midget impingers mounted at a height of 1.5 meters along five arcs. In addition, limited vertical sampling was carried out along the 100-meter arc by means of impingers mounted at nine levels on six lightweight towers. Electrically operated vacuum units, suitably positioned within the sampling network, provided aspiration for the impingers. During the diffusion experiments, air was drawn into the impingers through short sections of capillary tubing and bubbled through a dilute hydrogen-peroxide solution. Sulfur dioxide present in the air samples combined with the hydrogen peroxide to form sulfuric acid. Ten-minute averaged SO₂ concentrations were determined from laboratory measurements of the electrical conductivity of the aspired solutions. The samplers were arranged at 2-degree intervals along the 50, 100, 200, and 400-meter arcs (91 samplers per arc), and at 1-degree intervals for the 800-meter arc (i.e., 181 samplers along the 800-meter arc).

The meteorological conditions (wind speed and direction) were measured at a number of weather stations. Figure 1-4 shows the setup and typical concentrations recorded at the *Prairie Grass* field trials. Surface weather observations at a 2-meter height were recorded at the "Source," "North450," "MIT 3L," and "MIT AREA" weather stations. The "A&M TWR" surface tower provided wind speed and direction at seven levels (0.3, 0.5, 1.0, 2.0, 4.0, 8.0, and 16.0 meters) above the surface. In addition, the "Rawinsond" radiosonde was used to provide meteorological conditions at upper levels.

Approximately 70 experiments were carried out in a wide variety of weather conditions. Some of the trials (30, 31, 63, and 64) did not include correction factors for the compensation of the evaporative loss of impinger solution during aspiration.⁶ Other *Prairie Grass* trials have significant portions of meteorological data missing. For these

⁵ See References 1-7 and 1-8.

⁶ See Reference 1-7, pages 200-201.

reasons, 19 trials were removed from this comparison. Reference 1-1, Appendix B gives further details for the elimination of the *Prairie Grass* trials. Table 1-1 lists trials that were used in this study.

On some trials, critical information associated with a specific sampler arc was missing. The 50-meter arc of trial 62 did not include a correction factor for the compensation of the evaporative loss. The 200-meter arc of trial 50 and the 100-meter arc of trial 57 had vacuum lines disconnected.⁷ We eliminated these three arcs from our comparisons.

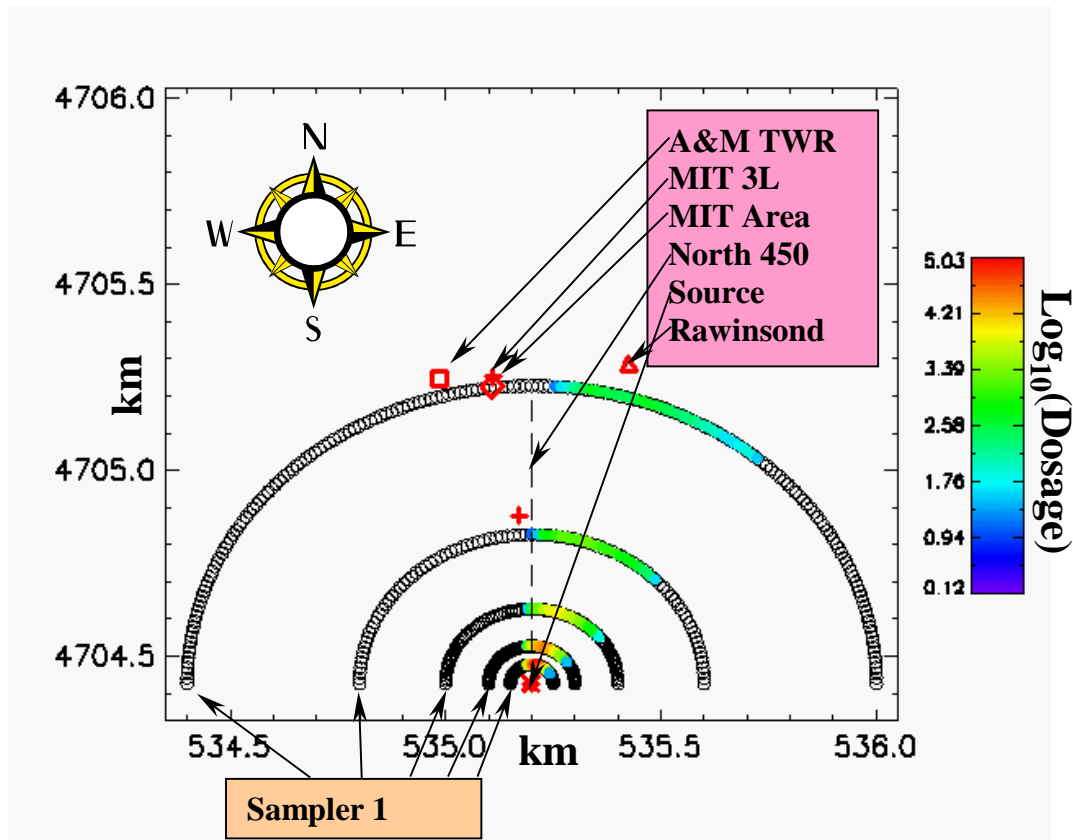


Figure 1-4. Example Dosages (in mg-sec/m³) and Field Trial Setup⁸

⁷ See Reference 1-7, pages 79-80.

⁸ The dosages shown here were obtained from *Prairie Grass* field trial 61. The color intensity corresponds to the logarithm of dosage at the samplers. Dosages varied over five orders of magnitude for this trial (thus logarithms are shown). Dosages are calculated by multiplying ten-minute averaged SO₂ concentrations, as reported in the *Prairie Grass* field trials, by 600 seconds. The source release point is at (535.2 km, 4704.427 km). The sampling network is oriented on a west-east grid, with the middle sampler pointing toward the north. Sampler 1 corresponds to the west-most sampler. Red symbols denote locations of the meteorological stations.

Table 1-1. List of *Prairie Grass* Trials That Were Used

5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 54, 55, 56, 57, 58, 59, 60, 61, 62

Computer-ready data files of the *Prairie Grass* field trials were provided by Lawrence Livermore National Laboratory [Ref. 1-9]. Appendix K of Reference 1-1 provides a list of data anomalies that we discovered during our analysis of the *Prairie Grass* field trial data and the “corrections” that we applied.

Because of the extensive data collection (both sampling and meteorological) and detailed analytical efforts associated with *Prairie Grass*, these field trials represent an excellent opportunity for detailed model prediction comparisons – arguably a short-range baseline. However, there are several practical limitations associated with any conclusions based only on model comparisons to these field trials. First, there are limitations associated with the source term. Only gaseous, short-term (i.e., 10-minute), single point source releases were considered. That is, particles and aerosols, which may be representative of certain types of releases were not examined. Similarly, line sources, moving sources, multiple sources, and instantaneous releases were not examined. The short-range (i.e., 800 meters) and flat terrain associated with all of the *Prairie Grass* field trials do not allow for the exploration of complex weather or terrain features. In addition, data associated with instantaneous concentration and plume time-of-arrival were not collected during these field trials. Therefore, comparisons of predicted instantaneous concentrations, which might be important for assessments of certain sensors, cannot be accomplished with only these field trials.

REFERENCES

- 1-1. Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T, and Larson, D., *User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models*, IDA Paper P-3554, January 2001.
- 1-2. Warner S., Heagy, J. F., Platt, N., Larson, D., Sugiyama, G., Nasstrom, J. S., Foster, K. T., Bradley, S., and Bieberbach, G., *Evaluation of Transport and Dispersion Models: An Initial Comparison of Hazard Prediction and Assessment Capability (HPAC) and National Atmospheric Release Advisory Center (NARAC) Predictions*, IDA Paper P-3555, May 2001.
- 1-3. Warner, S., Carpenter, J. N., Cook, J. M., Miller, R. S., and Hegemann, B. E., *NBC Hazard Prediction Model Capability Analysis*, IDA Document D-2245, September 1999.
- 1-4. DTRA, *The HPAC User's Guide: Version 3.2*, October 1999.
- 1-5. Bradley, S., Mazzola, T., Ross, R., Srinivasa, D., Fry, R., and Bacon, D., *Verification and Validation of HPAC 3.0*, for Defense Special Weapons Agency, June 1998, and references therein; Sykes, R. I., "HPAC/SCIPUFF: Kamisiyah Modeling Issues," *3rd Annual GMU/DTRA Transport and Dispersion Modeling Workshop*, Fairfax, VA, 28-29 July 1999; and Nappo, C. J., Eckman, R. M., Shankar Rao, K., Herwehe, J. A., and Gunter, L., *Second Order Closure Integrated Puff (SCIPUFF) Model Verification and Evaluation Study*, Air Resources Laboratory, NOAA, May 1998.
- 1-6. Sykes, R. I., et. al., *PC-SCIPUFF Version 1.0 Technical Documentation (DRAFT)*, A.R.A.P. Report 717, Titan Research and Technology Division, Princeton, pp. 103-109, 1998.
- 1-7. Barad, M. L. (Editor), *Project Prairie Grass, A Field Program in Diffusion*, Geophysical Research Papers No. 59, Volumes I and II, DTIC #AD-152572/AFCRC-TR-58-235(I), Air Geophysical Laboratory, Hanscom Air Force Base, MA, 1958
- 1-8. Haugen, D. A. (Editor), (1959) *Project Prairie Grass, A Field Program in Diffusion*, Geophysical Research Paper, No. 59, Volume III, Report AFCRC-TR-58-235, Air Force Cambridge Research Center.
- 1-9. The sampler data in the arcs were obtained from Doug Murray of TRC, while the 100-meter arc tower data were transcribed by LLNL (private communication, July 2000).

CHAPTER 2

METHODOLOGIES

2. METHODOLOGIES

This chapter provides a description of the methodologies used to compare HPAC probabilistic prediction outputs to the *Prairie Grass* field trial observations. First, various methods of computing 2D MOE values are identified. Next, this chapter describes the preparation of HPAC probabilistic predictions and provides some discussion of the *Prairie Grass* field trials. Finally, the quantitative communication of a user's potential risk tolerance is discussed in terms of the two-dimensional MOE space.

A. COMPUTATION OF 2D MOE¹

This section describes methods for computing the various regions, false positive (A_{FP}), false negative (A_{FN}), and overlap (A_{OV}), associated with the 2D MOE. Often, field trial observations are represented by dosages at specific sampler locations, perhaps along an arc or line. Therefore, a direct estimate of the area that contains a given concentration or dosage is not available. When using the above MOE for comparisons to observations, an approach to estimating the “area” size and location must be developed.

1. MOE Based on a Threshold

One can compute analogs to the area values, A_{FP} , A_{FN} , and A_{OV} , based on a critical threshold dosage. For sampler data collected along an arc, as was the case for the *Prairie Grass* field trials, this can be done directly. First, a threshold dosage of interest is chosen. One then can, for example, for the false positive region, compute the distance covered by the samplers in which there is a prediction but no observation above the threshold. Similarly, distances for the overlap region and false negative region can be computed. Figure 2-1 illustrates this procedure with a threshold of 60 mg-sec/m³ for the 800-meter arc samplers of *Prairie Grass* field trial 8. These *distances* are naturally related to actual areas. That is, this estimate represents a natural analog to a hazard area at a given critical threshold.

¹ Much of this section is extracted from Reference 2-1. Furthermore, additional information associated with the computation and interpretation of this MOE can be found in Reference 2-1.

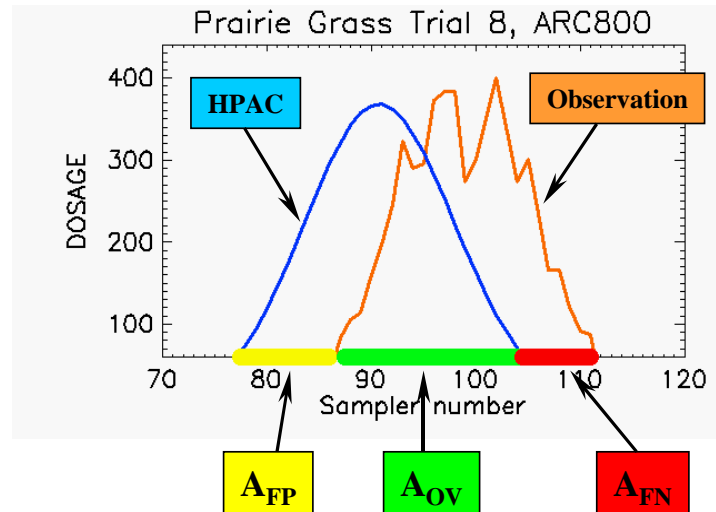


Figure 2-1. Illustration of Computations Based on a Threshold (Summed Distances Along the Arc)

2. MOE Based on Summed Dosages

Figure 2-2 presents an alternative scheme. Here the false positive region is really the dosage predicted in a region but not observed. Therefore, for A_{FP} (as shown in Figure 2-2), one first considers all of the samplers at which the *prediction is of greater value than the observation*. Next, one sums the differences between the predicted and observed dosages at those samplers. Based on the samplers that contained *observed values that were larger than the predicted values*, one can similarly compute A_{FN} . A_{OV} is calculated by considering all samplers and summing the dosages associated with the minimum predicted or observed value. These estimates can be made on a linear scale, as shown in Figure 2-2, or on a logarithmic scale.

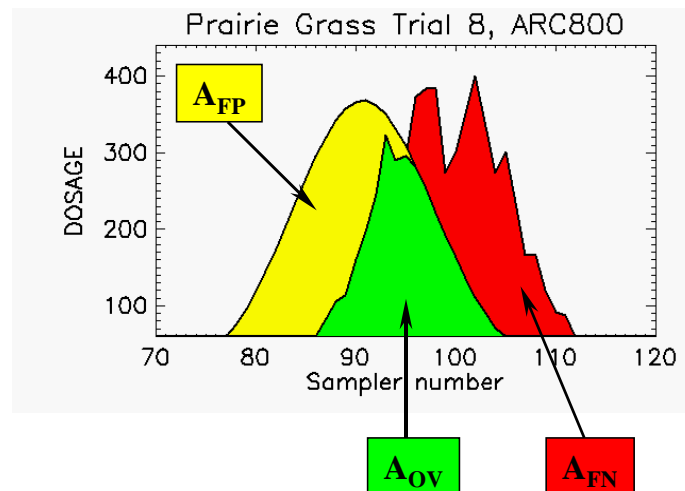


Figure 2-2. Illustration of Computations Based on Summed Dosages

One can also interpret the areas associated with observations and predictions by considering the lethality or effects of a presumed agent. This “lethality/effects filter” methodology is described in Reference 2-1, but is not used in this study.

3. MOE Based on an Interpolated Area

Each of the techniques described above – threshold-based, summed dosage-based, and lethality/effects-filtered – can be used to compute MOE values directly from the sampler data. One might also like to compute actual areas, for instance, to describe the impact on a nominal affected population. In this section, we describe two interpolation techniques that we used to generate physical “hazard” areas associated with the field trial data. With these interpolated regions, MOE values based on a critical threshold (e.g., a hazard area²), summed dosages, or a lethality/effects filter could again be computed. For the relatively short-range, densely sampled *Prairie Grass* field trial, results based on interpolation are not expected to differ much from those based on the procedures described above.

For the *Prairie Grass* field trials, measurements were taken at discrete samplers along five semi-circular arcs as shown in Figure 2-3.

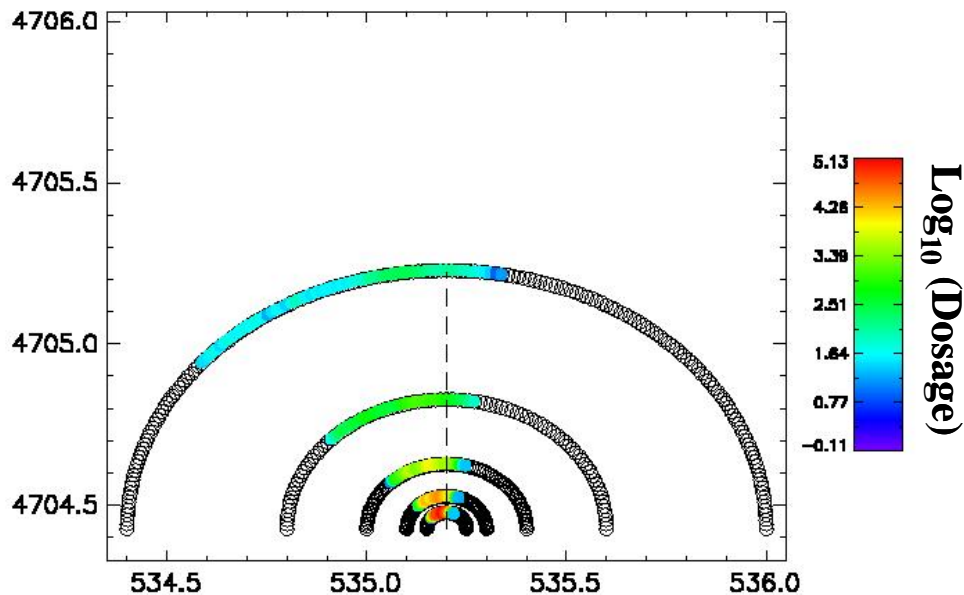


Figure 2-3. Example Dosages (in mg-sec/m³) for *Prairie Grass* Field Trial 43

² The hazard area is defined as the area enclosed within some critical (“hazard”) threshold contour.

Given values at some discrete (and irregular) set of samplers, the process of interpolation provides intermediate values on some regular grid of points. The resulting regular grid of functional values could be used to obtain contours of “hazard” areas (areas within a critical threshold contour) or calculate MOE values based on interpolated areas. Interpolated values can also be used to display “hazard” areas according to dosage intensity as is shown in Figure 2-4.³

Interpolation procedures can be carried out either in linear or logarithmic space. When interpolating actual plume dosages varying over orders of magnitude, we employ interpolation schemes in logarithmic space. When interpolating probabilistic plumes (where plume values are probabilities in [0,1]), we employ interpolation schemes in both linear and logarithmic spaces and present both results.

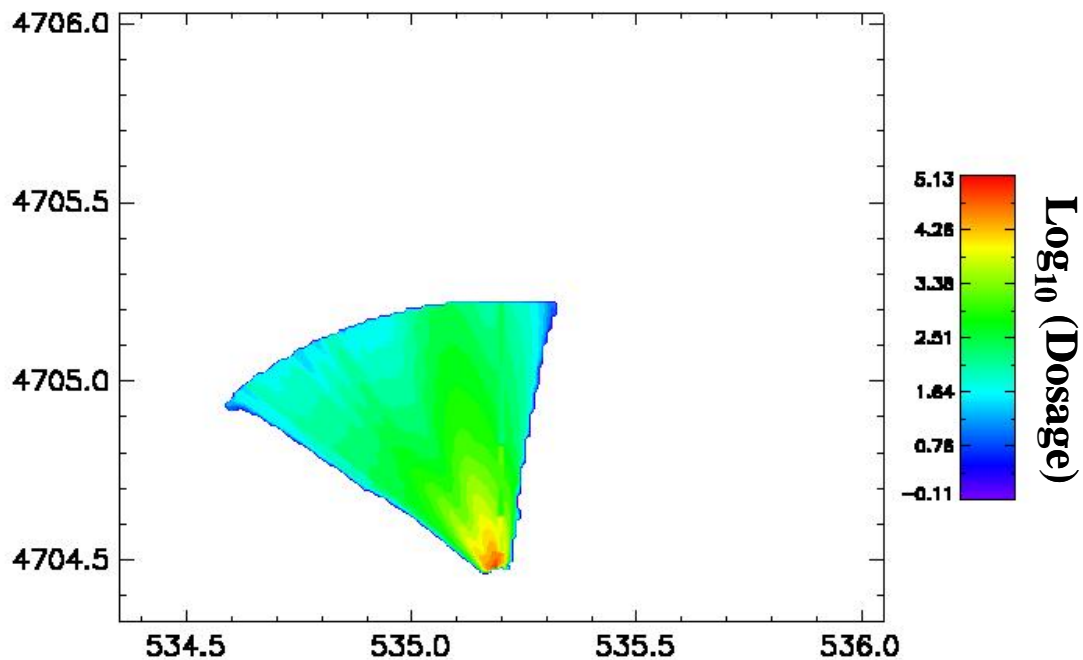


Figure 2-4. Example Dosages (in mg-sec/m³) for *Prairie Grass* Field Trial 43 Using Delaunay Triangulation Procedure to Perform Interpolation to a Regular Grid of Points

³ A complementary way to look at an “interpolated area” procedure is that it is one of many “average-weighting” techniques that assign a set of weights to a discrete set of samplers. For that purpose, an interpolated area assigns a natural weighting to a set of samplers that is particularly useful to distinguish samplers that are “clustered” from a “solitary” sampler. We plan to examine a few interpolation procedures and techniques in detail in a future study.

Two interpolation procedures, labeled “Polar Bilinear” and “Delaunay Triangulation” are described below. The *Prairie Grass* experimental setup was particularly amenable to these two procedures. Both interpolation routines were applied with a resolution of 1 m^2 resulting in 1601×801 grid points.

a. Polar Bilinear

Recalling that field trial samplers were arranged on semi-circular arcs at 2-degree intervals along the 50, 100, 200, and 400 meter arcs (91 samplers per arc) and at 1-degree intervals for the 800-meters arc (181 samplers along the 800-meter arc), we realize that *Prairie Grass* field trials are ideal candidates for a polar bilinear interpolation scheme. The polar bilinear scheme is employed in the polar domain and is analogous to bilinear interpolation employed in a rectangular domain [Ref. 2-2]. Consider the situation depicted in Figure 2-5 [Ref. 2-3].

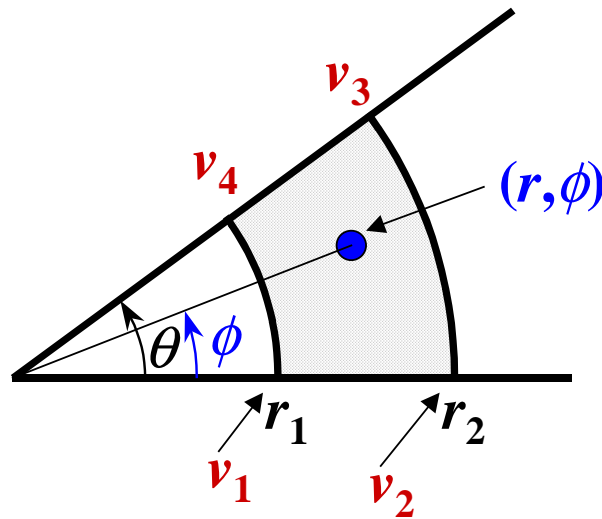


Figure 2-5. Illustration of Polar Bilinear Interpolation

Given the “corner” values v_1, v_2, v_3, v_4 at coordinates $(r_1, 0), (r_2, 0), (r_2, \theta), (r_1, \theta)$, respectively, we desire an interpolated value at coordinates (r, ϕ) inside the shaded region (i.e., the region contained within $(r_1, 0), (r_2, 0), (r_2, \theta), (r_1, \theta)$). Here θ represents the angular sampler separation, and ϕ represents an angular “offset” into the interpolating semi-annulus (shaded region). The polar bilinear interpolation function that gives this value should have the following two properties. First, for any constant radius r , $r_1 \leq r \leq r_2$, the interpolated value should vary linearly between the values at $(r, 0)$ and (r, θ) as the polar angle ϕ , $0 \leq \phi \leq \theta$, is varied. Similarly, for any constant polar angle ϕ , $0 \leq \phi \leq \theta$, the interpolated value should vary linearly between the values at (r_1, ϕ) , and

(r_2, ϕ) as the radius r , $r_1 \leq r \leq r_2$, is varied. The following function has these properties [Ref. 2-3]:

$$f(r, \phi) = \left(\frac{r - r_1}{r_2 - r_1} \right) \left(\frac{\phi}{\theta} \right) v_3 + \left(\frac{r - r_1}{r_2 - r_1} \right) \left(\frac{\theta - \phi}{\theta} \right) v_2 + \left(\frac{r_2 - r}{r_2 - r_1} \right) \left(\frac{\phi}{\theta} \right) v_4 + \left(\frac{r_2 - r}{r_2 - r_1} \right) \left(\frac{\theta - \phi}{\theta} \right) v_1. \quad (2-1)$$

For the semi-annular regions contained between the 50- and 100-, 100- and 200-, and 200- and 400-meter arcs, the above procedure could be applied directly. For the semi-annular region associated with the 400- and 800-meter arcs, we first linearly interpolated the 2-degree inter-sampler interval of the 400-meter arc to a 1-degree interval, and then applied the above procedure.

b. Delaunay Triangulation

The Delaunay triangulation procedure is useful for the interpolation, analysis, and visual display of irregularly, discretely gridded data. From a set of discrete points (sampler coordinates), a planar triangulation is formed, satisfying the property that the circumscribed circle of any triangle in the triangulation contains no other vertices in its interior.⁴ For any point that is within some triangle (formed via Delaunay triangulation), a linear interpolation routine using values at the vertices of the triangle is used to calculate the value at that point. Delaunay triangulation is efficiently implemented in IDL⁵ and forms a core interpolation routine for display of irregularly gridded data.

Within IDL, the area interpolation procedure is accomplished by calls to the TRIANGULATE procedure to obtain Delaunay triangulation of the sampler locations followed by the TRIGRID procedure that performs linear interpolation of sampler values to a regular grid.

Figure 2-4 shows an application of this methodology to *Prairie Grass* field trial 43.

4. Estimation of Confidence Intervals

In order to assess the significance of a particular result, we computed confidence regions (95 and 80 percent) for MOE point estimates. We used bootstrap (resampling) procedures to estimate these confidence regions [Ref. 2-6]. First, for each trial (and each individual arc), an MOE value, (x,y), was computed. For a given set of comparisons, for example, all 50-meter arcs, the vector mean of these values corresponded to the MOE

⁴ Delaunay triangulation is the dual structure of the Voronoi diagram [Ref. 2-4].

⁵ IDL = Interactive Data Language [Ref. 2-5].

point estimate. Approximate confidence regions were computed by resampling MOE vectors. For example, when examining the 10 neutral stability trials, 1,000 bootstrap samples of size 10 were formed. From these, 1,000 mean vectors (i.e., bootstrapped MOE values) were computed. The 950 bootstrapped MOE values that were closest (as measured by Euclidean distance) to the actual MOE point estimate constituted our estimated 95th percent confidence region.

B. PREPARATION OF HPAC PROBABILISTIC PREDICTION OUTPUTS

This section provides information on the HPAC output that was considered in this study. The source term and meteorological inputs used for the 51 *Prairie Grass* field trial predictions in this study are identical to those previously used for comparison of HPAC (mean value) and NARAC outputs [Ref. 2-1]. For the present analysis, we use HPAC “.tab” file outputs to obtain probabilistic and mean value results. Our previous analysis of HPAC *Prairie Grass* field trial predictions considered only mean value outputs and used HPAC “.smp” file outputs.⁶ With respect to MOE results, the usage of these different formats, .tab and .smp, led to minor differences between mean value outputs. However, some issues with HPAC plotting/data output software were uncovered. Appendix B of this paper provides additional details.

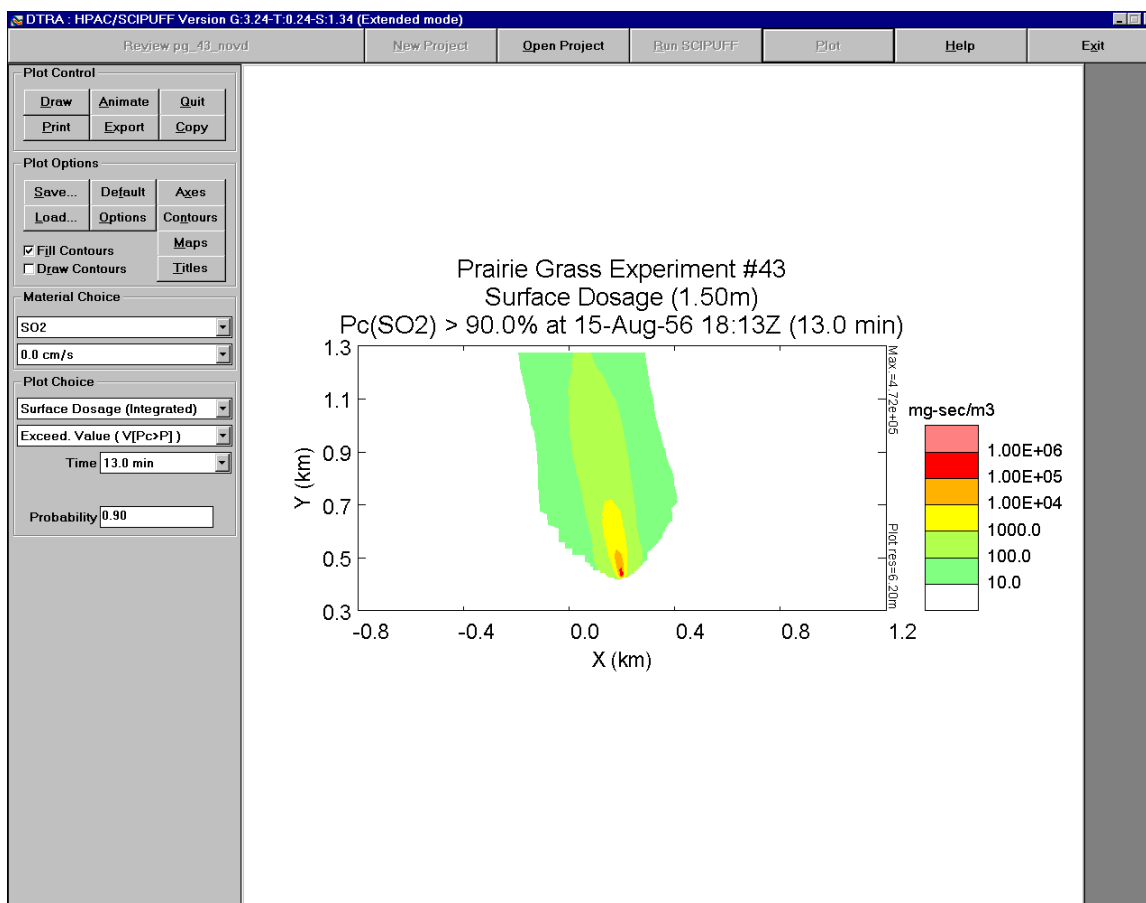
For this study, HPAC probabilistic outputs were examined. Two types of probabilistic outputs were investigated. First, “conditional probability” prediction outputs were studied. Figure 2-6, a “screen dump” from an HPAC prediction of *Prairie Grass* field trial 43, illustrates the plotted output associated with a conditional probability prediction. These outputs are created by toggling “Exceed. Value ($V[P_c > P]$)” on the HPAC plotting screen in the *Plot Choice* dialog box. The example shown in Figure 2-6 is for a conditional probability of 0.90 (i.e., $P = 0.90$).

The displayed areas obtained from the above procedure correspond to regions where the conditional probability of exceeding the contour value is greater than the selected value (0.90 in Figure 2-6.) For example, in Figure 2-6, the area inside the outermost (green) contour corresponds to the region in which the conditional probability of values exceeding 10 mg-sec/m³ is equal to or greater than 0.90. The conditional probability is defined as the probability given that the field value is greater than zero.⁷

⁶ HPAC predictions at specified sampler locations are output in the .smp file.

⁷ The mean value output is defined as the mean field value.

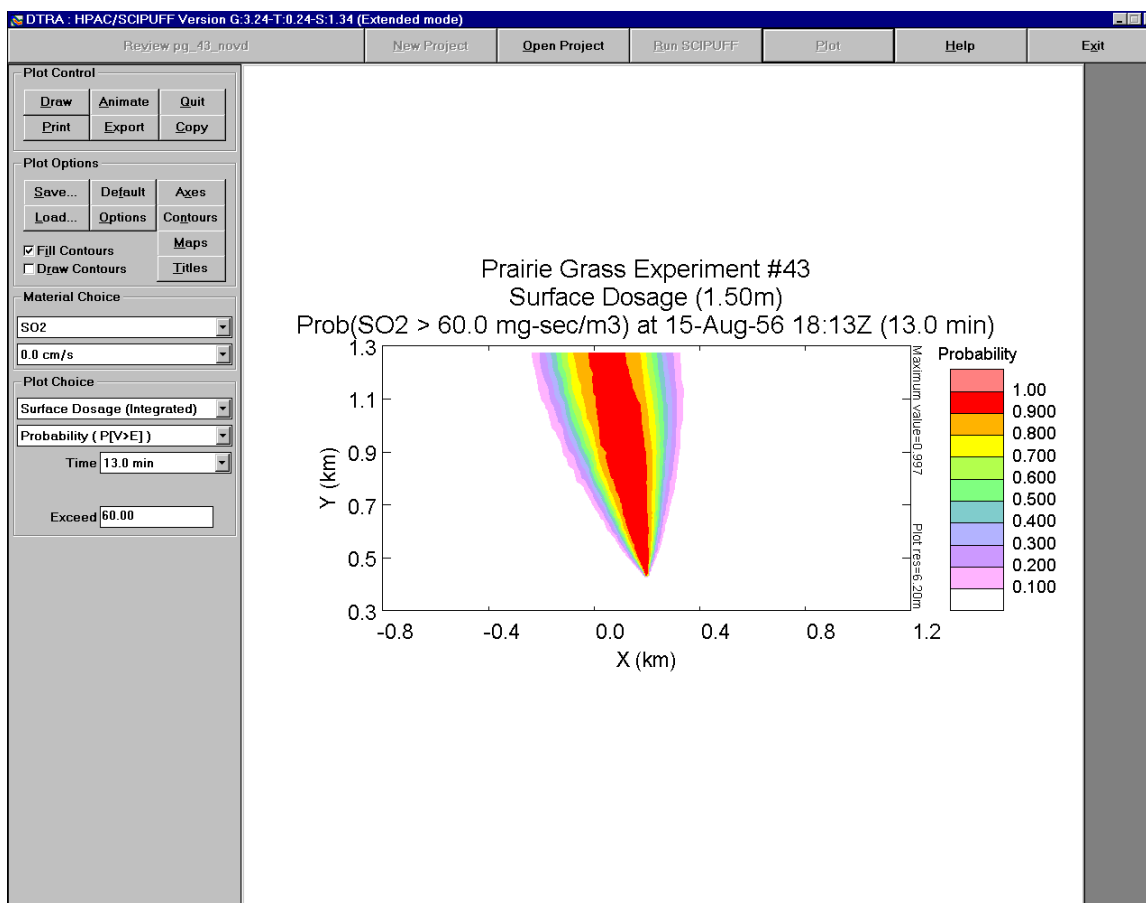
This measure neglects the possibility of intermittent zero dosages. It is intended to be a *conservative* estimate of the risk of exposure [Ref. 2-7].



**Figure 2-6. Illustration of Conditional Probability Prediction Output:
"Exceed. Value (V[P_c > 0.90])"**

The conditional probability output provided in the HPAC .tab file contains dosages at each defined location. We defined the locations in the .tab file to coincide with the *Prairie Grass* sampler locations.

Figure 2-7 provides a screen dump from an HPAC prediction that illustrates the plotted output for a "full probability" prediction. These outputs are created by toggling "Probability (P [V>E])" on the HPAC plotting screen in the *Plot Choice* dialog box. The example shown in Figure 2-7 is based on a threshold value of 60 mg-sec/m³. Figure 2-7 shows contours of the probability that the selected field will exceed a prescribed value (in this case, 60 mg-sec/m³). For example, the outermost contour (light purple) corresponds to a probability of 0.10 of exceeding 60 mg-sec/m³.



**Figure 2-7. Illustration of Full Probability Prediction Output:
“Probability (P [V > 60 mg-sec/m³])”**

The full probability output provided in the HPAC .tab file contains probabilities, from 0 to 1.0, at each defined location. We defined the locations in the .tab file to coincide with the *Prairie Grass* sampler locations. Therefore, when considering full probability outputs, assessments of MOE values at a critical threshold value can be done directly. For example, at the full probability prediction value of 0.10 and a threshold of 60 mg-sec/m³, any sampler that falls within the outermost contour in Figure 2-7 would be considered to have a prediction above threshold, therefore contributing to either the false positive (A_{FP}) or overlap (A_{OV}) region, depending on the field trial observations, when computing the MOE.

MOEs based on summed dosages cannot be directly computed from the HPAC full probability output described above. Therefore, while MOEs based on a threshold value of 60 mg-sec/m³ and summed dosages were computed for conditional probability

prediction outputs, only threshold-based MOE values were computed for full probability outputs.

Appendix C contains conditional probability prediction outputs compared to *Prairie Grass* field trials. Conditional probabilities of 0.01, 0.50, 0.80, 0.85, 0.90, 0.95, and 0.999 and the mean value output are shown as a function of arc range for each of the 51 trials in Appendix C. Full probability prediction outputs of 0.01, 0.50, 0.80, 0.85, 0.90, 0.95, and 0.999 were also examined in this study. Appendix D compares the full probability prediction output for a probability of 0.50, the 0.50 conditional probability output, and the *Prairie Grass* field trial data. Again, the five arcs for each of the 51 trials are presented.

C. PRAIRIE GRASS DATA

HPAC predictions of 51 *Prairie Grass* field trials were used for this study.⁸

1. Structure of Field Trial

Each *Prairie Grass* trial contains data for five semicircular arcs of samplers placed at an altitude of 1.5 meters, and six towers containing samplers at various altitudes located at the 100-meter arc [Ref. 2-8]. In this study, only samplers located at semicircular arcs at the height of 1.5 meters were considered. Figure 2-8 depicts a typical *Prairie Grass* trial result for the five arcs that were used in the analyses.

Table 2-1 lists the number of samplers in each of the semicircular arcs, angular separation between samplers in the arc, and the Cartesian distance between samplers in the arc. For each arc, a single value⁹ for the user-oriented MOE is calculated, thus yielding five separate values per trial. These values are combined into separate groupings that are then used in the comparisons. This structure of the *Prairie Grass* field trials gives a natural way of comparing results by arc range.

⁸ A total of 70 trials were associated with the *Prairie Grass* experiment. Critical measurements were missing or known to be inaccurate for some trials. Therefore, not all trials were used for this study. See Reference 2-1, Appendices B and K, for additional details related to the selection of these 51 trials.

⁹ For the two-dimensional MOE, this is represented by a vector (x,y).

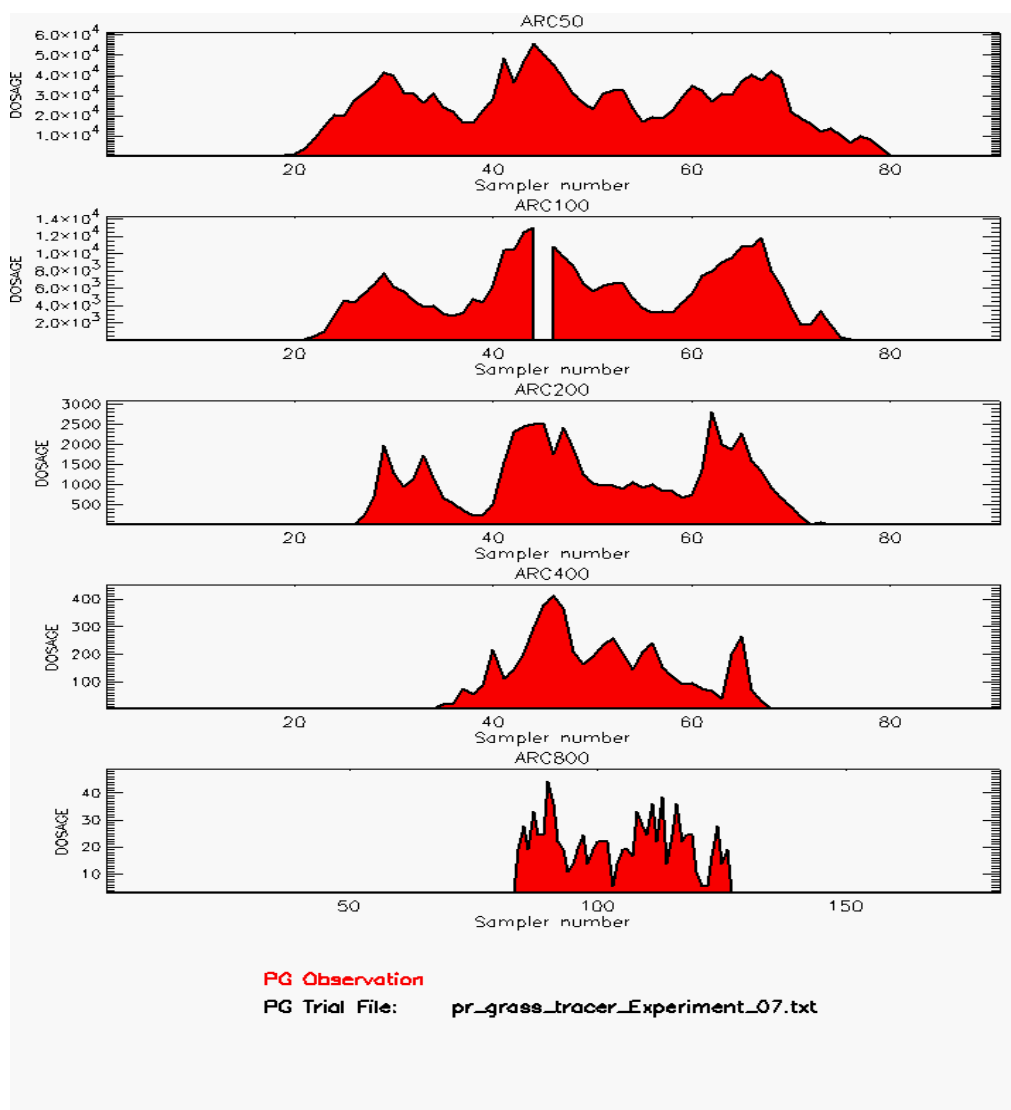


Figure 2-8. *Prairie Grass* Trial 7 as a Function of Arc: Stability Category is 1¹⁰

Table 2-1. Distance Between Samplers by Arc

Arc	Number of Samplers	Angular Separation, degrees	Cartesian Separation, meters
50-meter	91	2	1.75
100-meter	91	2	3.49
200-meter	91	2	6.98
400-meter	91	2	13.96
800-meter	181	1	13.96

¹⁰ The value for sampler 45 of the 100-meter arc is missing. Dosage units are in mg-sec/m³. Dosage times a breathing rate (in m³/sec) equals dose (in mg).

Thus, there is a natural way to combine MOEs calculated for each individual arc into a single MOE value that is associated with a particular trial. To do so, we associate each sampler with a short line segment centered at the sampler location and having the same length as the inter-sampler distance for that arc. Then for some specified threshold, we multiply the number of samplers in each individual arc used to define A_{OV} , A_{FN} , and A_{FP} by the inter-sampler distance. Adding the values for the five arcs together forms the A_{OV} , A_{FN} , and A_{FP} estimates for the entire (all arcs) trial. Similarly, one can compute values based on summed dosages by combining the crosswind-integrated dosages for each individual arc together by multiplying them by the inter-sampler distance for each individual arc.

2. Stability Category Groupings

Stability category assignments, which were developed by Irwin and Rosu [Ref. 2-9] for the *Prairie Grass* trials, were used to group trials with similar characteristics in terms of atmospheric stability. There are seven stability categories ranging from “very unstable” (category 1) to “very stable” (category 7). Because of the relatively small sample sizes involved, we used two different procedures to combine categories together, as depicted in Table 2-2. Comparing results obtained by these categorizations of the data allowed for the examination of the statistical spread to assess the effects of the relatively small sampling sizes for some groupings. Table 2-3 lists the *Prairie Grass* data sample sizes that were used in the model comparisons.

Table 2-2. Stability Category Groupings (SCGs) Used in This Study^a

	SCG Set 1	SCG Set 2
Unstable	1,2	1,2,3
Neutral	3,4,5	4
Stable	6,7	5,6,7

^a A more physically-based arrangement would be to group the near neutral categories, 3 and 4, together.

Table 2-3. Sample Sizes Used for Model Comparisons to *Prairie Grass* Field Trials^a

SCG	All Arcs	50-Meter	100-Meter	200-meter	400-meter	800-meter
1	6	6	6	6	6	4 ^b
2	11	10 ^c	11	10 ^d	11	11
3	9	9	8 ^e	9	9	9
4	10	10	10	10	10	10
5	7	7	7	7	7	7
6	6	6	6	6	6	5 ^f
7	2	2	2	2	2	2
Total	51	50	50	50	51	48
First procedure used to combine stability categories						
1-2	17	16	17	16	17	15
3-4-5	26	26	25	26	26	26
6-7	8	8	8	8	8	7
Second procedure used to combine stability categories						
1-2-3	26	25	25	25	26	24
4	10	10	10	10	10	10
5-6-7	15	15	15	15	15	14

- ^a Sub-groupings for the comparisons include all, by range, by all arcs and stability group, and by range and stability group.
- ^b The 800-meter arcs of trials 7 and 16 have maximum values below the data threshold of 60 mg-sec/m³.
- ^c The 50-meter arc of trial 62 has a missing correction factor for the compensation of the evaporative loss of the impinger solution during aspiration [Ref. 2-8, p. 201].
- ^d The 200-meter arc of trial 50 was replaced by “missing values” because the vacuum line to sampler 62 became disconnected during the run [Ref. 2-8, p. 80].
- ^e The 100-meter arc of trial 57 was replaced by “missing values” because the vacuum line to sampler 47 was believed to have become disconnected [Ref. 2-8, p. 80].
- ^f The 800-meter arc of trial 39 has too many samplers that are missing. Of 19 samplers that recorded non-zero values, 13 have missing values. For this case, we replaced the whole arc with “missing values.”

3. Threshold for Sampler Observations

SO₂ concentrations were estimated in the following way. First, SO₂ was captured at the impingers (samplers). These impingers were later filled with a hydrogen peroxide (H₂O₂) solution. This caused the SO₂ and H₂O₂ to form sulfuric acid (H₂SO₄). Next, the

concentration of sulfuric acid was estimated by measuring the electrical conductivity. Based on the computed sulfuric acid concentration, an estimate of the original SO₂ concentration was derived.

The reported sampler minimal averaged concentration value for the *Prairie Grass* field trials is 0.005 mg/m³, which corresponds to a minimal dosage of 3 mg-sec/m³. However, to obtain averaged concentration values for samplers inside the gas plume, it was necessary to estimate the “background” averaged concentrations for SO₂. To do so, some samplers were placed outside of the limits of the time-mean gas plume. In general, this small background concentration level was almost entirely due to a trace amount of sulfuric acid that had been added during the preparation of the dilute hydrogen peroxide solution. Because of this, as the limit of the sampling technique was approached, the uncertainty of the averaged SO₂ concentration determination increased rapidly. For concentrations less than 0.1 mg/m³ (60 mg-sec/m³), this uncertainty was estimated to be 25 percent.¹¹

The mechanism described above was considered the chief source of uncertainty associated with estimates of SO₂ concentrations and, hence, dosages. For this reason, we chose 0.1 mg/m³ (60 mg-sec/m³) as the data cutoff threshold.

D. QUANTITATIVE COMMUNICATION OF RISK TOLERANCE: “USER-COLORING”

We now turn our attention toward acceptable risk tolerance. The user of a transport and dispersion model may typically be faced with the tradeoff between false negative and false positive. In this section, we describe two notional scoring functions that a user might employ in order to communicate and analyze his or her application-specific risk tolerance preferences.

1. Risk-Weighted Figure-of-Merit (R-FOM)

Typical operational users of transport and dispersion models might consider false positives and false negatives quite differently. For many applications, false positives would be much more acceptable to the user than false negatives (which could result in decisions that directly lead to death or injury). Equation 2-2, described in Reference 2-1, is an example of a possible user scoring function. Basically, Equation 2-2 describes a

¹¹ See Reference 2-8, p. 77.

spatial figure-of-merit (FOM) that evaluates the ratio of overlap area to total (predicted plus observed) area, for example, above a specified threshold.¹² The coefficients, C_{FN} and C_{FP} , weight the false negative and false positive regions, respectively. Hence, we refer to this notional user scoring function as the Risk-Weighted FOM (R-FOM).

$$R-FOM = \frac{A_{OV}}{(A_{OV} + C_{FN}A_{FN} + C_{FP}A_{FP})} \quad (2-2)$$

where $C_{FN}, C_{FP} > 0$.

It may be true that, for some applications (e.g., technical model validation), the weightings for false negatives and false positives are considered irrelevant or set equal ($C_{FN} = C_{FP}$).¹³

Equation 2-3 relates the x- and y-axes of the 2D MOE to R-FOM. This equation is derived by first recognizing that for the 2D MOE, $x = \frac{A_{OV}}{(A_{OV} + A_{FN})}$, $y = \frac{A_{OV}}{(A_{OV} + A_{FP})}$ and, therefore, $A_{FN} = \left(\frac{1-x}{x}\right)A_{OV}$ and $A_{FP} = \left(\frac{1-y}{y}\right)A_{OV}$. These definitions of A_{FN} and A_{FP} are then substituted into Equation 2-2. Following algebraic rearrangement, Equation 2-3 is obtained.

$$R-FOM = \frac{xy}{(xy + C_{FN}y(1-x) + C_{FP}x(1-y))} \quad (2-3)$$

Figure 2-9 shows MOE contours (i.e., isolines for R-FOM) for $C_{FN} = C_{FP} = 1$. Similarly, Figure 2-10 illustrates the case where A_{FN} is weighted by a factor of 10 relative to A_{FP} and 5 relative to A_{OV} (i.e., $C_{FN} = 5$ and $C_{FP} = 0.5$).

¹² Reference 2-10 describes a related spatial FOM that does not include explicit weighting coefficients.

¹³ As developed here, the implicit coefficient associated with A_{OV} is 1.0. Therefore, the notion of equal weights for A_{FN} and A_{FP} (i.e., $C_{FN} = C_{FP}$) is insufficient for the complete specification of R-FOM. That is, the precise R-FOM values will depend on the values chosen for C_{FN} and C_{FP} and not just their ratio.

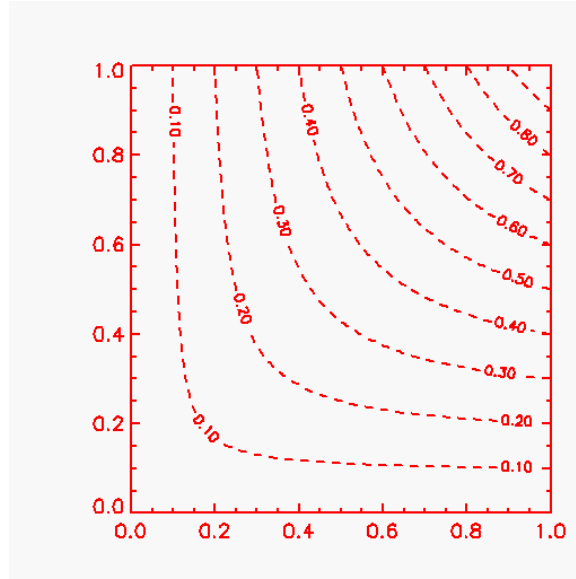


Figure 2-9. Relationship Between R-FOM and 2D MOE: $C_{FN} = 1$, $C_{FP} = 1$

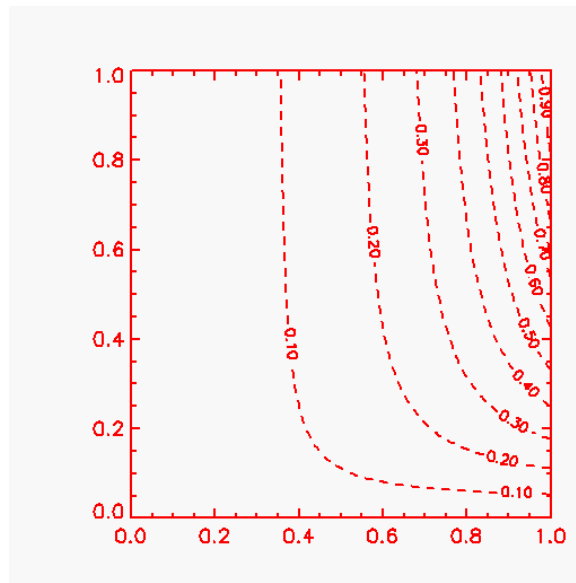


Figure 2-10. Relationship Between R-FOM and 2D MOE: $C_{FN} = 5$, $C_{FP} = 0.5$

The two figures above show that estimates of the 2D MOE can be directly converted to estimates of R-FOM. For example, a 2D MOE point estimate of (0.5,0.5) implies, according to Figure 2-9 (and Equation 2-3), an R-FOM value of 0.33 for $C_{FN} = C_{FP} = 1.0$. Similarly, inspection of Figure 2-10 shows that for $C_{FN} = 5$ and $C_{FP} = 0.5$, a 2D MOE point estimate of (0.5,0.5) suggests an R-FOM value of about 0.15.

The contours above (i.e., isolines) can be used as the basis for coloring the MOE space. Figure 2-11 provides example user coloring of the MOE space based on R-FOM.

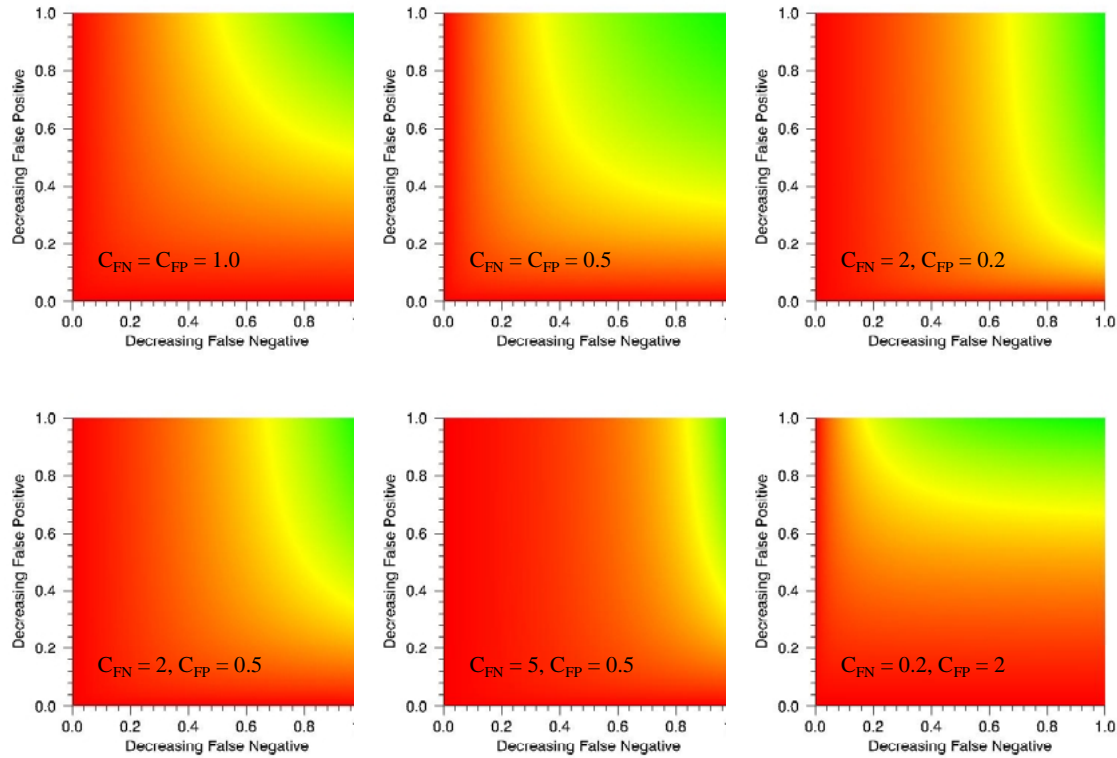


Figure 2-11. User Coloring of MOE Space: R-FOM

Figure 2-12 quantifies the coloring scheme used for the plots shown in Figure 2-11. The x-axes shown in Figure 2-12 show the color indices from 1 to 256, from “pure” red to “pure green.” The bottom portion of the figure shows these gradations. The y-axis, running from 0 to 255, denotes the color intensity.

This plot can be “read” in the following way. The color intensity of “blue” is always 0 (i.e., no blue added). At an R-FOM of 0.0, this coloring scheme incorporates pure red. Importantly, as the user-defined FOM (e.g., R-FOM) increases from 0.0 to 0.50, the intensity of green increases linearly. For instance, at an R-FOM value of 0.5, there are equal intensities of red and green (hence, yellow). Similarly, for R-FOM values between 0.50 and 1.0, the red intensity is reduced linearly with increasing R-FOM value. At an R-FOM (or A-FOM) value of 1.0, the coloring used is pure green. The bar at the

bottom of Figure 2-12 illustrates RGB color additions (from the top of the figure). This bar also associates the resulting colors with MOE values.

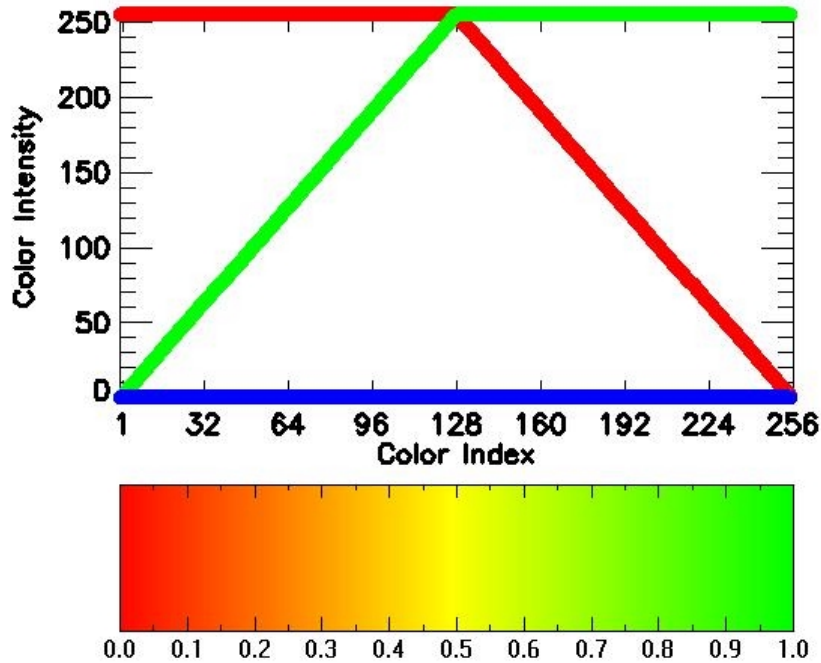


Figure 2-12. Illustration of Red-Green-Blue (RGB) Coloring Scheme

2. Area Size-Only FOM (A-FOM)

A transport and prediction model might be applied in such a way that the actual location of the hazard or direction of the plume was of no particular importance. For example, such a model might be used to study potential future outcomes of an accidental or intentional release. In this case, the actual weather (e.g., wind speed and direction) of the future (for instance one year in the future) cannot be known with any accuracy. This researcher may wish a scoring function that simply compares the sizes of the predicted and observed areas.¹⁴

To begin development of this FOM, we recall that the 2D MOE is defined by

$$(x, y) = \left(\frac{A_{OV}}{A_{OB}}, \frac{A_{OV}}{A_{PR}} \right). \quad (2-4)$$

¹⁴ Here, we use a generalized version of the phrase “sizes of the areas.” This notion could include physical areas within specified hazard contours, marginal dosages associated with false positive and false negative “areas,” or the associated incremental/marginal casualties.

Next, let us consider the ratio

$$\frac{y}{x} = \frac{A_{OV}}{A_{PR}} / \frac{A_{OV}}{A_{OB}} = \frac{A_{OB}}{A_{PR}} . \quad (2-5)$$

Next, consider points in the 2D MOE space that lie on the diagonal line, that is, the line $y = x$. Then,

$$1 = \frac{y}{x} = \frac{A_{OB}}{A_{PR}} \Rightarrow A_{PR} = A_{OB} . \quad (2-6)$$

Thus, this diagonal in the 2D MOE space consists of the points that incorporate “equal size” predictions and observations. Let us assume a hypothetical requirement that A_{PR} and A_{OB} must be within a factor of s of each other, with $s > 1$. Mathematically, this is stated by requiring that

$$\frac{1}{s} \leq \frac{A_{PR}}{A_{OB}} \leq s . \quad (2-7)$$

Figure 2-13 shows A-FOM isolines, in MOE space, for various values of parameter s .

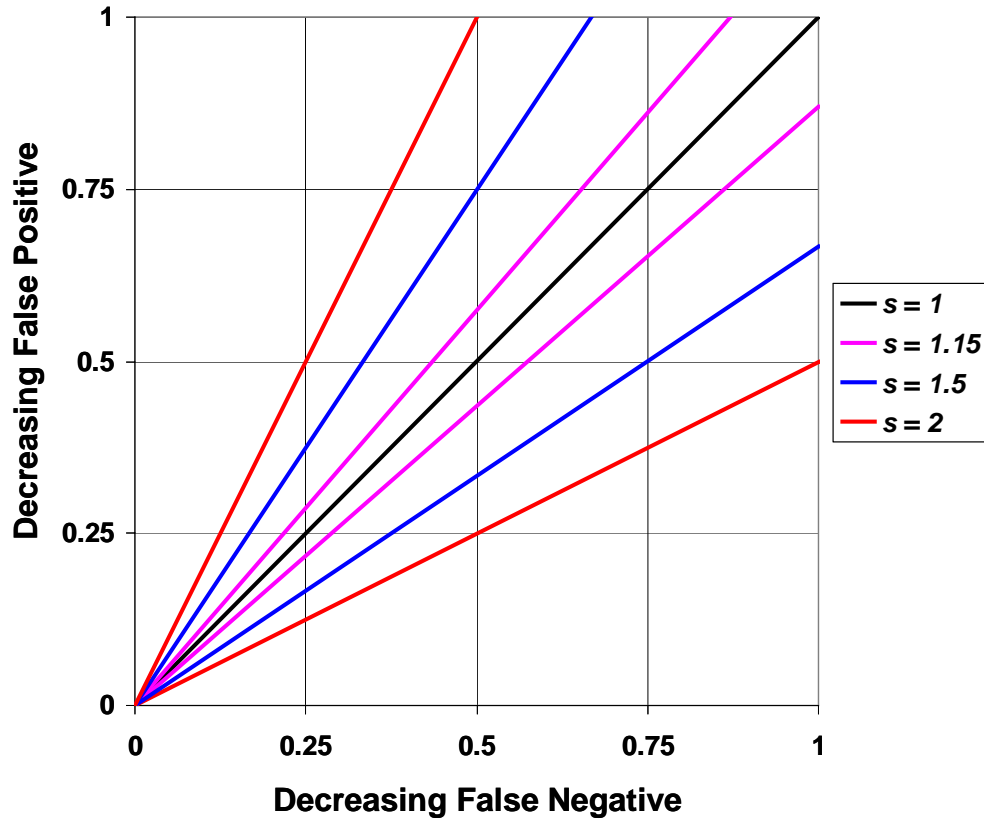


Figure 2-13. A-FOM Isolines for Some Values of the Parameter s

For any value of $s > 1$, bounding rays $\frac{A_{PR}}{A_{OB}} = s$ and $\frac{A_{PR}}{A_{OB}} = \frac{1}{s}$ separate the MOE space into two regions. Any MOE value within the bounding rays satisfies Equation 2-7, while any 2D MOE value outside the bounding rays implies that A_{OB} and A_{PR} are different by more than a factor of s .¹⁵

To construct a user coloring of the 2D MOE space using this concept we transform Equation 2-7 into polar coordinates as shown below.

$$\frac{1}{s} \leq \tan(\theta) \leq s \Rightarrow \begin{cases} \theta \leq \text{atan}(s) \\ \theta \geq \text{atan}\left(\frac{1}{s}\right) \end{cases} \quad (2-8)$$

Examining Figure 2-14, we note that the bounding rays for any given $s > 1$ are symmetric around the diagonal line $\theta = \pi/4$; thus we only need to define a coloring function on the $[0, \pi/4]$ domain and then extend it to the $[\pi/4, \pi/2]$ domain by symmetry. Let $s > 1$ be given. We would like the “Area Size-Only FOM,” A-FOM, to vary between 0 and 1 with the value of 0.5 on the bounding rays $\theta = \text{atan}(s)$ and $\text{atan}(1/s)$. Thus, we are looking for a function f satisfying the following properties:

$$\begin{cases} f(0) = 0 \\ f\left(\text{atan}\left(\frac{1}{s}\right)\right) = 0.5 \\ f(\pi/4) = 1 \end{cases} \quad (2-9)$$

Consider the function

$$f(\theta) = A\theta e^{B\theta}. \quad (2-10)$$

Substituting into Equation 2-9 and solving for the parameters A and B yields the following:

¹⁵ Note that Equation 2-5 could be rewritten as $\begin{cases} \frac{A_{PR}}{A_{OB}} \leq s \\ \frac{A_{OB}}{A_{PR}} \leq s \end{cases}$ Hence, for any given s , it is possible to state

that the predictions and observations are within $(s-1)*100$ percent of each other. For example, setting $s = 1.15$, Equation 2-5 implies that A_{PR} and A_{OB} are within 15 percent of each other. This symmetry between the predictions and observations is especially important for model versus model comparisons when either truth (e.g., field trial) data are unavailable or an *a priori* judgment cannot be made about the quality of either model’s predictions.

$$\left\{ \begin{array}{l} B = \frac{\ln \left(\frac{8 \operatorname{atan} \left(\frac{1}{s} \right)}{\pi} \right)}{\frac{\pi}{4} - \operatorname{atan} \left(\frac{1}{s} \right)} \\ A = \frac{4}{\pi e^{B\pi/4}} \end{array} \right. \quad (2-11)$$

Figure 2-14 shows a graph of this function, f , for $s = 1.15$.

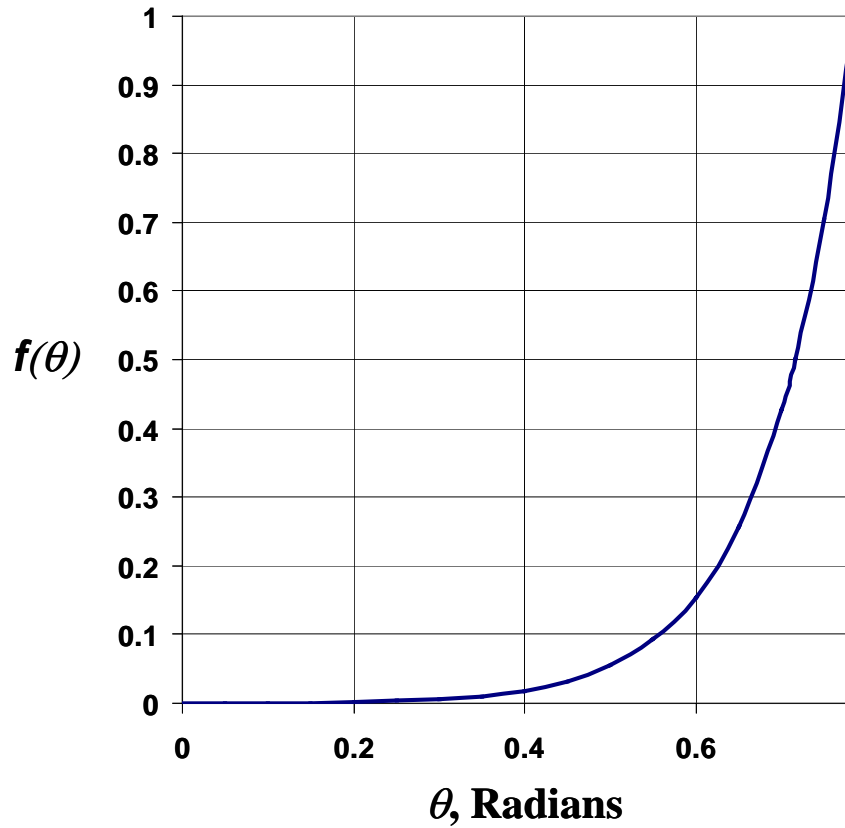


Figure 2-14. Graph of Function f Defined on $[0, \pi/4]$

Now, we use symmetry around the $\theta = \pi/4$ line to define this user coloring function in the polar domain as

$$\text{A-FOM}(\theta) = \begin{cases} f(\theta) & \text{if } 0 \leq \theta \leq \pi/4 \\ f(\pi/2 - \theta) & \text{if } \pi/4 \leq \theta \leq \pi/2 \end{cases} \quad (2-12)$$

Figure 2-15 shows a graph of this A-FOM as a function of the polar angle θ .

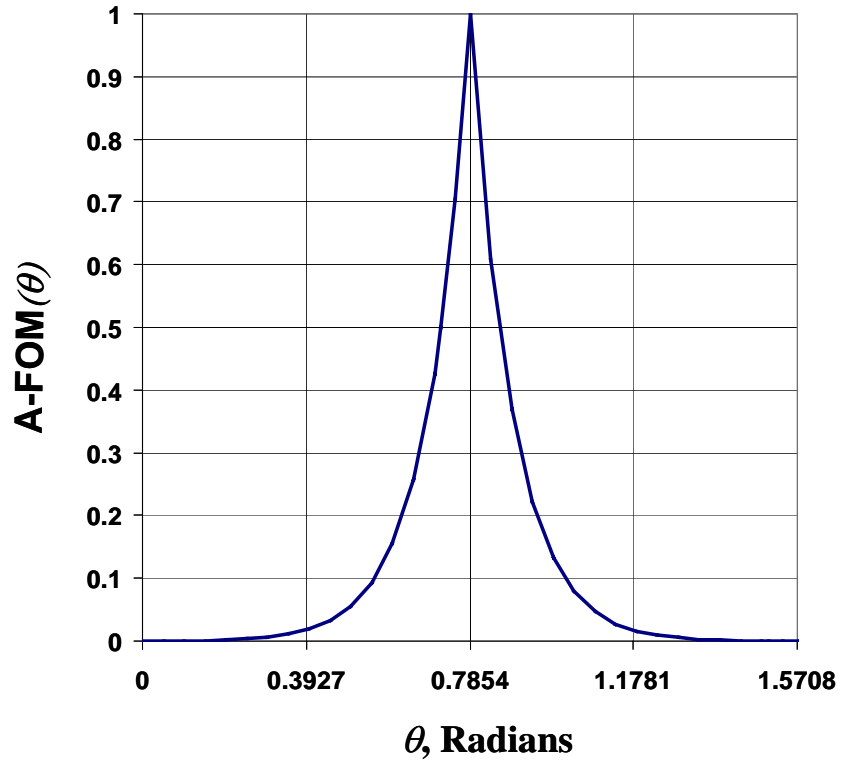


Figure 2-15. Graph of A-FOM Function Defined on $[0, \pi/2]$

Figure 2-16 shows user colorings of 2D MOE space based on A-FOM functions. The previously described linear RGB scheme was applied to these plots. For example, the bright yellow lines correspond to an A-FOM value of 0.5.

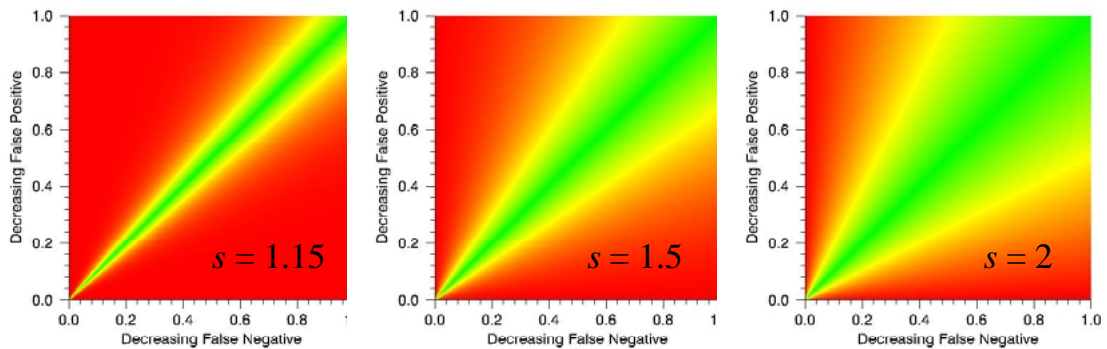


Figure 2-16. Examples of A-FOM User-Coloring for $s = 1.15, 1.5$ and 2

User colorings based on R-FOM and A-FOM will be used in Chapter 3 to describe notional MOE evaluations of HPAC probabilistic prediction outputs.

REFERENCES

- 2-1. Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T, and Larson, D., *User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models*, IDA Paper P-3554, January 2001.
- 2-2. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, Second Edition, Cambridge University Press, 1992.
- 2-3. Lowell Bruce Anderson, IDA, unpublished note, 2001.
- 2-4. Guibas, L. J., Knuth, D. E., and Sharir, M., "Randomized Incremental Construction of Delaunay and Voronoi Diagrams," *Algorithmica* 7: 381-413, 1992. Also, see <http://www.gris.uni-tuebingen.de/gris/proj/dt/dteng.html>.
- 2-5. *Interactive Data Language (IDL)* developed by Research Systems Inc., www.rsinc.com.
- 2-6. Efron, B. and Tibshirani, R. J., "An Introduction to the Bootstrap," *Monographs on Statistics and Applied Probability* 57, Chapman and Hall, 1993.
- 2-7. *HPAC/SCIPUFF Version G:3.232*, Help Command: Plot Choice, DTRA, October 1999.
- 2-8. Barad, M. L. (Editor), *Project Prairie Grass, A Field Program in Diffusion*, Geophysical Research Papers No. 59, Volumes I and II, DTIC #AD-152572/AFCRC-TR-58-235(I), Air Geophysical Laboratory, Hanscom Air Force Base, MA, 1958.
- 2-9. Irwin, J.S., and Rosu, M-R., "Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models," *Proceedings of the 10th Joint Conference on the Applications of Air Pollution Meteorology*, American Meteorological Society, Boston, pp. 6-10, 1998.
- 2-10. Mosca, S., Graziani, G., Klug, W., Bellasio, R., and Bianconi, R., "A Statistical Methodology for the Evaluation of Long-Range Dispersion Models: An Application to the ETEX Exercise," *Atmos. Environ.*, Vol. 32, No. 24, 4307-4324, 1998.

CHAPTER 3

RESULTS, ANALYSES, AND DISCUSSION

3. RESULTS, ANALYSES, AND DISCUSSION

This Chapter demonstrates the application of the MOE to the notional assessment of the “goodness” of HPAC probabilistic predictions of the short-range *Prairie Grass* field trials. HPAC full and conditional probabilistic predictions are compared for a critical threshold of 60 mg-sec/m.³ The computed MOE estimates support the notion that, by adjusting the HPAC input probability values, one can, in general, select a wide range of associated false positive and false negative fractions. Given a specific requirement, defined here in terms of user-coloring (R-FOM and A-FOM), one can select HPAC probability values that satisfy this user.

This chapter provides the results and discussion associated with this analysis. First, we discuss the performance of HPAC probabilistic prediction outputs as described by the 2D MOE. We present results for both HPAC “full” and “conditional” probabilities. MOE results associated with interpolated areas are also discussed in this Chapter. Next, evaluations of the computed values using notional user-colorings of the two-dimensional MOE space are presented. Finally, the concept of average affected area is introduced and described. This concept can be used to develop the absolute value, in terms of “hazard” area (in km²) or number of affected people, associated with a given MOE estimate. Future IDA analysis will likely take advantage of this property.

A. PROBABILISTIC OUTPUT RESULTS: PROBABILITY OF EXCEEDING A THRESHOLD

This section presents MOE results for HPAC full probability predictions.¹ All of the MOE estimates shown in this section are based on a threshold value of 60 mg-sec/m³.

1. Across All Trials

Figure 3-1 presents MOE 95th percent confidence regions for HPAC full probability predictions of the 51 *Prairie Grass* field trials. Also, the MOE region for the HPAC mean value prediction is shown in Figure 3-1 (the yellow region). It can be seen

¹ The full probability outputs were created by toggling “Probability (P [V>E])” on the HPAC plotting screen in the *Plot Choice* dialog box. See Chapter 2, Section B, for additional information.

that, in going from probability values of 0.999 to 0.01, the MOE indicates that the false negative decreases and the false positive increases. The MOE associated with the 0.50 probability prediction is closest to the value based on the HPAC mean value prediction for these data. With the exception of the 0.01 probability estimate, the chosen probabilistic values (0.50, 0.80, 0.85, 0.90, 0.95, and 0.999) led to increased false negative and decreased false positive fractions, relative to the mean value predictions.

We note that the predicted plume size (e.g., width) decreases with increasing probability value. For example, 0.01 full probability values are always associated with the widest predictions and the 0.999 full probability values are always associated with the narrowest predictions.² See Appendix C for comparative results

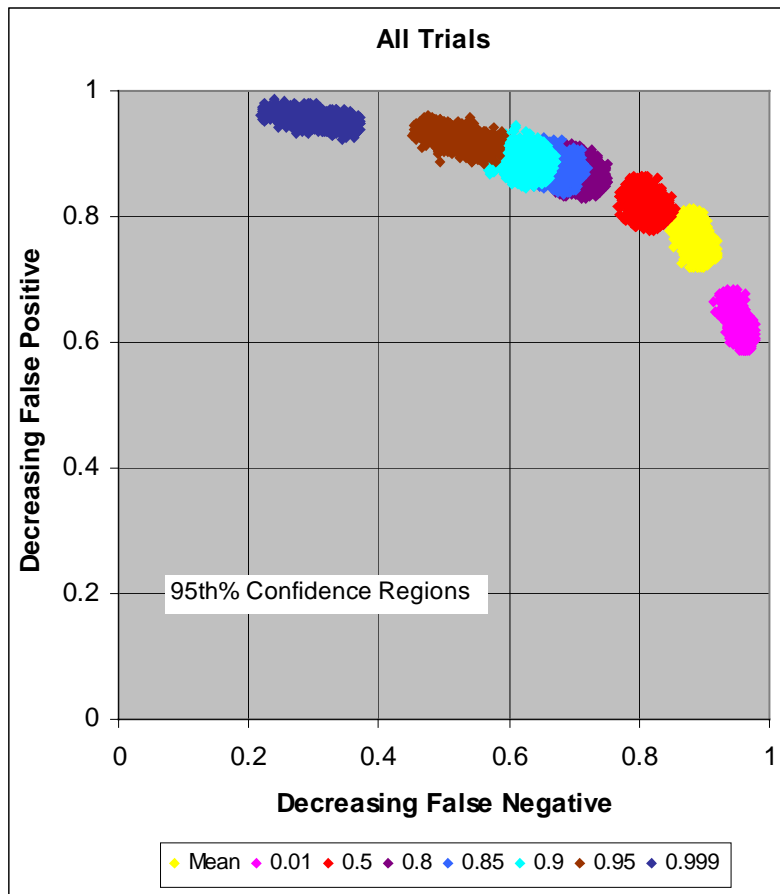


Figure 3-1. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full Probability Predictions: All Trials

² Alternatively, if we define a hazard area as wherever the dosage exceeds some critical threshold (e.g., 60 mg-sec/m³), then we can interpret the 0.01 full probability prediction as encompassing the area that has a 0.99 (1.00-0.01) probability of containing the hazard area.

2. As a Function of Range

Figure 3-2 presents results for all trials as a function of arc range (50, 100, 200, 400, and 800 meters). The overall MOE value degrades somewhat as the range increases. For example, inspection of the HPAC mean value prediction MOE confidence regions shows that both the false negative and false positive fractions increase, at least a small amount, as the range is increased from 50 to 800 meters.

For the 50- and 100-meter arc, there is essentially no false positive fraction for the probabilistic predictions, with the exception of the 0.01 probability predictions. This indicates that the HPAC probabilistic predictions at 50 and 100 meters are enveloped by the observations, at least when considering a threshold of 60 mg-sec/m^3 .

For the 50- through 200-meter arcs, the 0.01 probability prediction result is most like the mean value result. However, at 400 and 800 meters, the 0.50 probability predictions lead to MOE estimates that are closest to those of the mean value predictions. At 800 meters, the 0.50 probability prediction and mean value prediction MOE estimates are almost identical.

At all ranges, the MOE confidence regions based on the HPAC 0.01 probability predictions indicate the smallest false negative fraction (i.e., the most conservative prediction).

3. As a Function of Stability Category

Figure 3-3 provides MOE confidence regions as a function of meteorological stability category grouping. The basic trend for the MOE value for the different probabilistic outputs considered can again be seen for each of the SCGs.³

For the more unstable trials, $\text{SCG} = 1,2$ and $\text{SCG} = 1,2,3$, the MOE for the 0.01 probability predictions (magenta) indicates no false negative fraction, that is, the prediction envelops the observation, at least on average, at the threshold considered (60 mg-sec/m^3). Alternatively, the MOE based on the 0.999 probability prediction indicates that the observation completely envelops the prediction for the more unstable trials.

³ Trials were assigned stability categories based on a 1998 Irwin and Rosu analysis [Ref. 3-1].

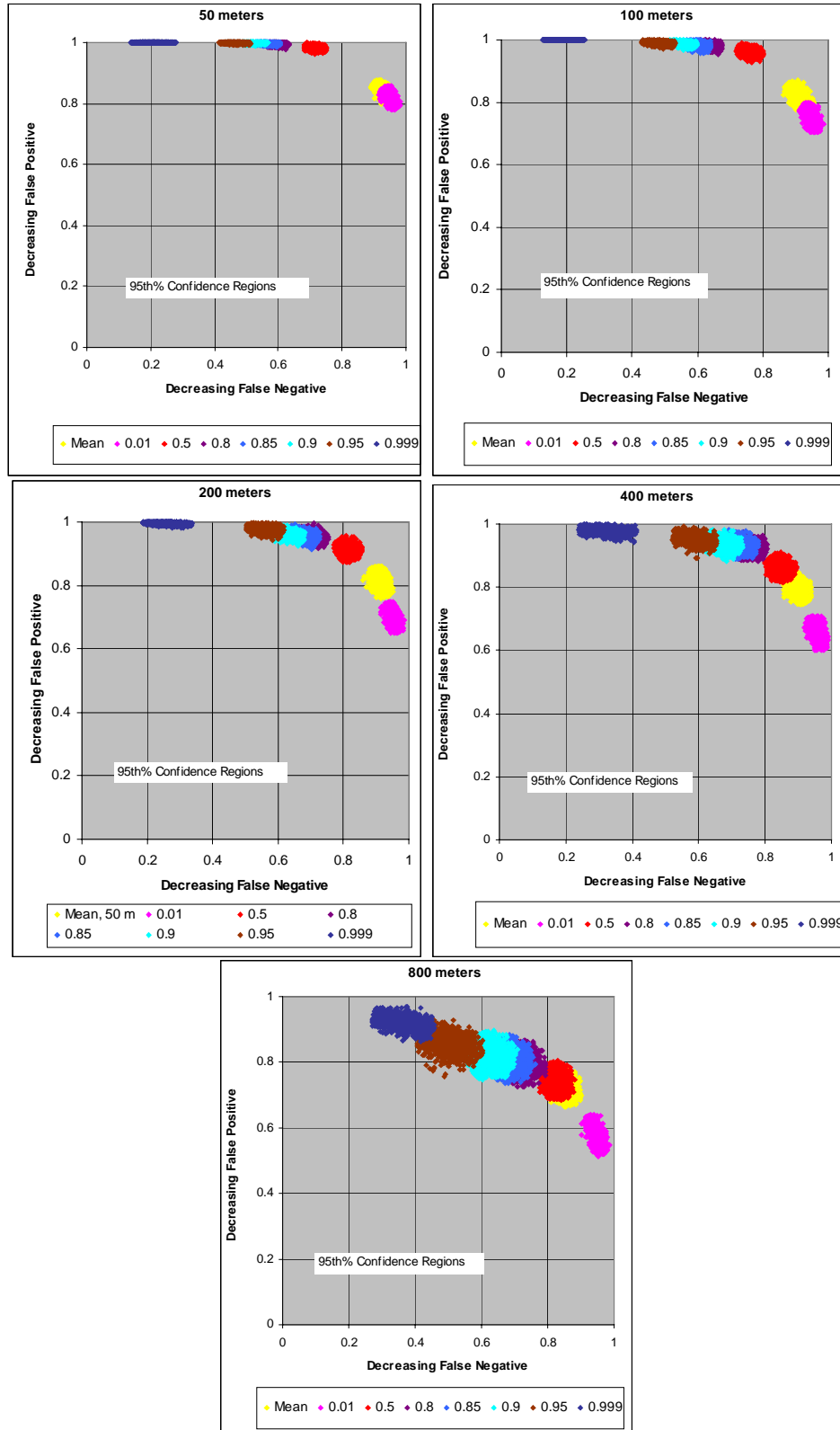


Figure 3-2. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full Probability Predictions: As a Function of Arc Range

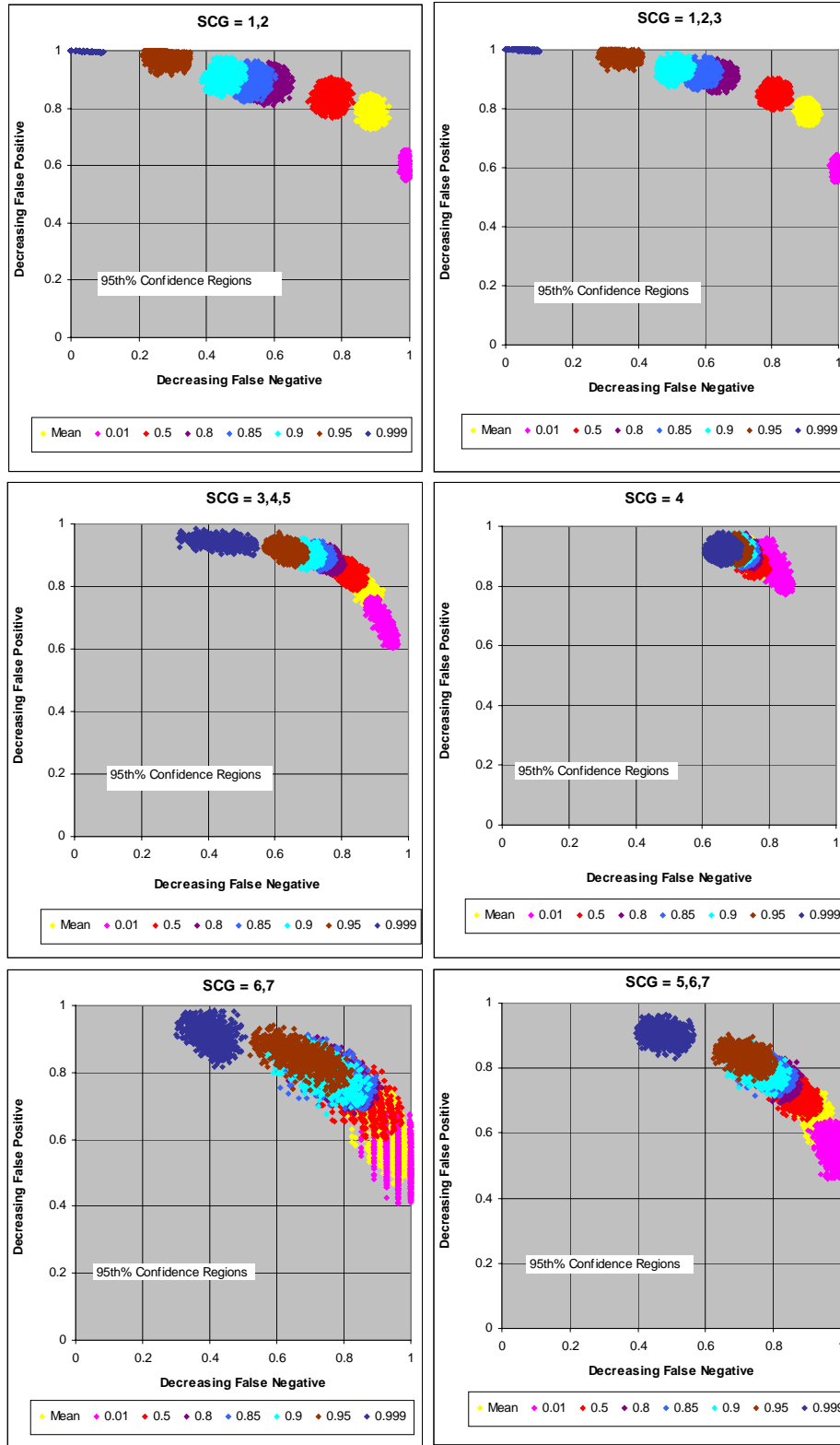


Figure 3-3. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full Probability Predictions: As a Function of Stability Category Grouping (SCG)

The MOE estimates for the neutral trials, SCG = 4, provide an interesting result. There is relatively little “motion” in the MOE values as a function of chosen probability value. This is in sharp contrast to results described thus far. For example, unlike the other SCGs, for the neutral trials, one cannot choose a reasonable probability value such that the false negative fraction is kept below 0.10. This relative lack of variability as a function of chosen probability may be a manifestation of the physical properties associated with neutral stability conditions. The suggestion is, that for these predictions, a user would not have much choice with respect to false negative/positive fractions for the neutral conditions.

It is plausible that this lack of variability as a function of selected probability value for the neutral trials is also related to the adopted protocol used for these short-range comparative predictions. For these predictions, the Monin-Obukhov length (L)⁴ was directly defined (“hard-wired”).⁵ This was done as part of the original HPAC/NARAC comparison protocol [Ref. 3-3] that was agreed upon. The L values were available for the *Prairie Grass* experiment because of the extensive data collection and subsequent analytical efforts associated with the field trial. For many model applications, the direct input of L values will not be possible.

For the most stable trials, SCG = 6,7, the large size of the 95th percent confidence regions is caused, in part, by the relatively small sample size, 8, for those trials.⁶

4. Analogy to Receiver Operating Characteristic (ROC) Curve

The curves that would be traversed by connecting the MOE values for the different probability levels (0.01 through 0.999 shown in Figures 3-1 through 3-3) satisfy several of the criteria associated with classic interpretations in signal detection theory. In signal detection theory, one considers a given stimulus (e.g., a signal to be detected in noise). For that stimulus, a given detector (e.g., a person or detection system) is said to operate along a specific Receiver Operating Characteristic (ROC) curve [Ref. 3-4]. For instance, a detection system can be set up, perhaps by adjusting a detection threshold, to

⁴ The Monin-Obukhov L value refers to a characteristic length scale involving the heat flux, buoyancy, and friction velocity that arises in the specification of turbulent diffusion near the earth’s surface [3-2].

⁵ Reference 3-3, Table B-1 (Appendix B) provides some of the parameters used to create the HPAC predictions. In particular, we note here that for the neutral trials (SCG =4), the Monin-Obukhov length scale (in meters) varied from 76 to 248.

⁶ Comparisons of MOE estimates based on HPAC mean value predictions as a function of SCG have been previously reported [Ref. 3-3].

achieve a certain probability of detection (P_d) against a given stimulus. Accordingly, some signals are missed, and thus there is a probability of missing a stimulus ($1-P_d$). For that set of conditions, a certain false alarm rate, that is, a probability of detecting a signal when there was no stimulus (P_{fa}), also exists. The ROC curve paradigm tells us that, by making some decision behavior change to the detector, one changes both the P_d (or $1-P_d$) and the false alarm rate together. For example, one might lower the signal detection threshold such that P_d is greatly increased (almost all signals are detected). However, the false alarm probability would necessarily increase. In fact, the ROC curve, for a given detection system and against a given stimulus, describes exactly how the probability of detection and false alarm vary as changes are made to the decision behavior of the detector. Figure 3-4 shows a ROC curve for some notional detector and stimulus.

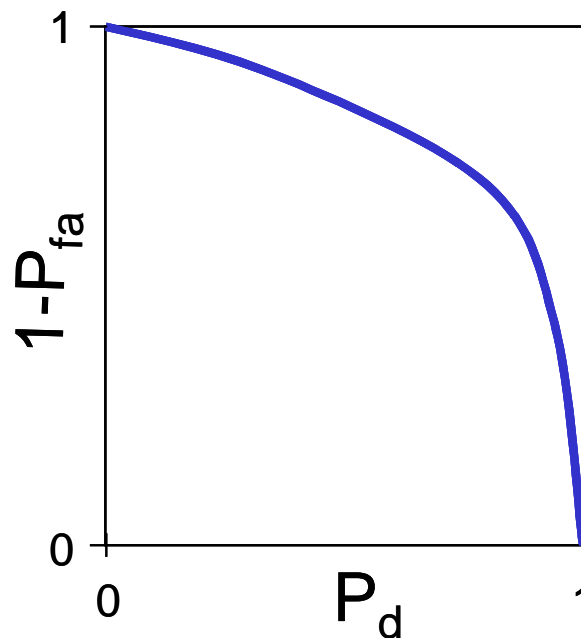


Figure 3-4. Notional ROC Curve: Trade-Off Between Detection and False Alarm Probabilities for a Given Sensor and Stimulus with Various Points on the Blue Line Corresponding to Different Decision Behavior

The analogy between the false negative/positive fractions and $1-P_d/P_{fa}$ appears to be reasonable. Thus invoking this ROC paradigm, we interpret the various MOE values for the different probability outputs (0.01 through 0.999) as simply different decision rules. That is, a user determines which decision rule (e.g., probability output) is most appropriate for his/her situation. For example, an application or user that demands a very low chance of inadvertent population exposure (conservative) might employ the 0.01 probability prediction outputs in order to minimize the false negative fraction.

B. PROBABILISTIC OUTPUT RESULTS: CONDITIONAL PROBABILITY

This section provides some results and discussion based on HPAC “conditional probability” prediction outputs.⁷ For the conditional probability predictions, MOE estimates based on a threshold of 60 mg-sec/m³ and on summed dosage were examined.

1. MOEs Based on Threshold of 60 mg-sec/m³

Figure 3-5 compares MOE confidence regions for HPAC full and conditional probability predictions. The results based on conditional probability predictions follow a pattern similar to those reported for the full probability predictions. However, inspection suggests that the full probability predictions lead to decreased false positive and increased false negative fractions.

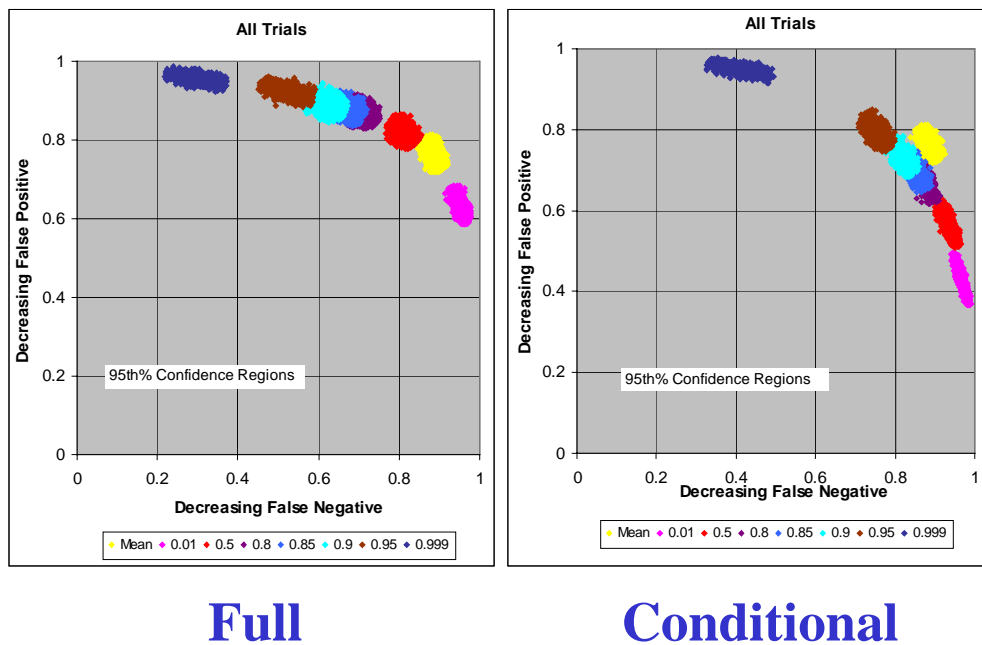


Figure 3-5. MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full and Conditional Probability Predictions: All Trials

Figure 3-6 directly compares the MOE confidence regions based on the two types of HPAC probability predictions, full and conditional. In Figure 3-6, the blue regions correspond to the six conditional probability results, the red regions correspond to the six full probability results, and the yellow region corresponds to the mean value result. The “best” predictions, that is, the predictions that led to MOE estimate closest to the perfect

⁷ These outputs were created by toggling “Exceed. Value (V[Pc>P])” on the HPAC plotting screen in the Plot Choice dialog box. See Chapter 2, Section B, for additional information.

(1,1) value, occur for both the 0.50 full probability and mean value predictions (which are approximately equidistant from the (1,1) point).

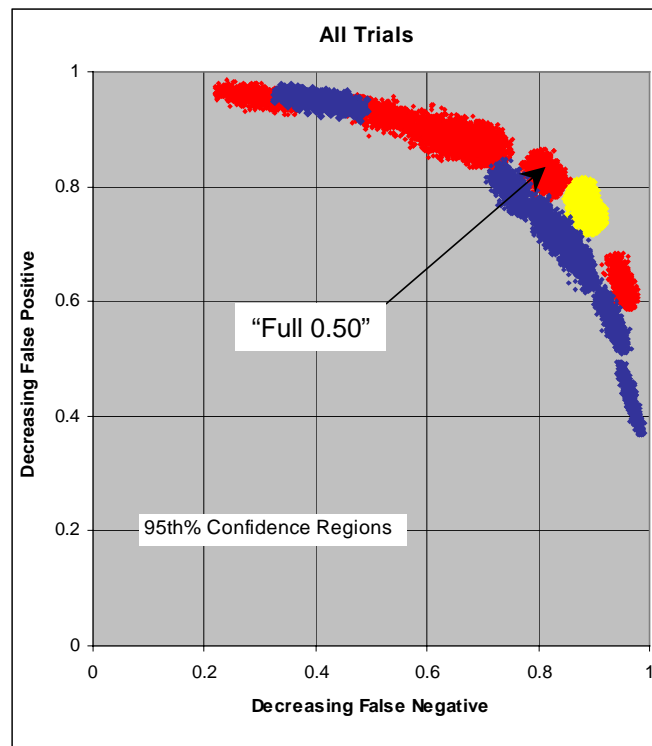


Figure 3-6. Direct Comparison of HPAC Full (Red Regions) and Conditional Probability (Blue Regions) Predictions (with Mean Value Prediction in Yellow): MOE Estimates (Based on 60 mg-sec/m³) for All Trials

The comparisons shown in Figure 3-6 suggest that the full probability predictions lead to narrower plumes (smaller false positive and larger false negative fractions) at a threshold of 60 mg-sec/m³ than do the corresponding conditional probability predictions. Figures 3-7 and 3-8 compare the 0.50 full probability predictions with the 0.50 conditional probability predictions for two *Prairie Grass* trials, a stable (Trial 58) and an unstable (Trail 16) trial, respectively.⁸

⁸ In Figures 3-7 and 3-8, the solid lines correspond to the observed values and the 0.50 conditional probability prediction (denoted “Exceed. Value (V[Pc>P])” on the HPAC plot screen in the Plot Choice dialog box). The meanings of the colors used in the plots are described in the below legend:

- Red: trial dosages that are higher than the 0.50 conditional probability prediction
- Green: trial dosages that overlap with the 0.50 conditional probability prediction
- Yellow: 0.50 conditional probability predicted dosages that are higher than the trial observations.

The blue diamonds denote samplers where HPAC predicted a 0.50 (or greater) probability of exceeding 60 mg-sec/m³. These outputs are denoted as “Probability (P [V>E])” on the HPAC plot screen in the Plot Choice dialog box. The 60 mg-sec/m³ dosage level is highlighted by dashed lines in the below figures.

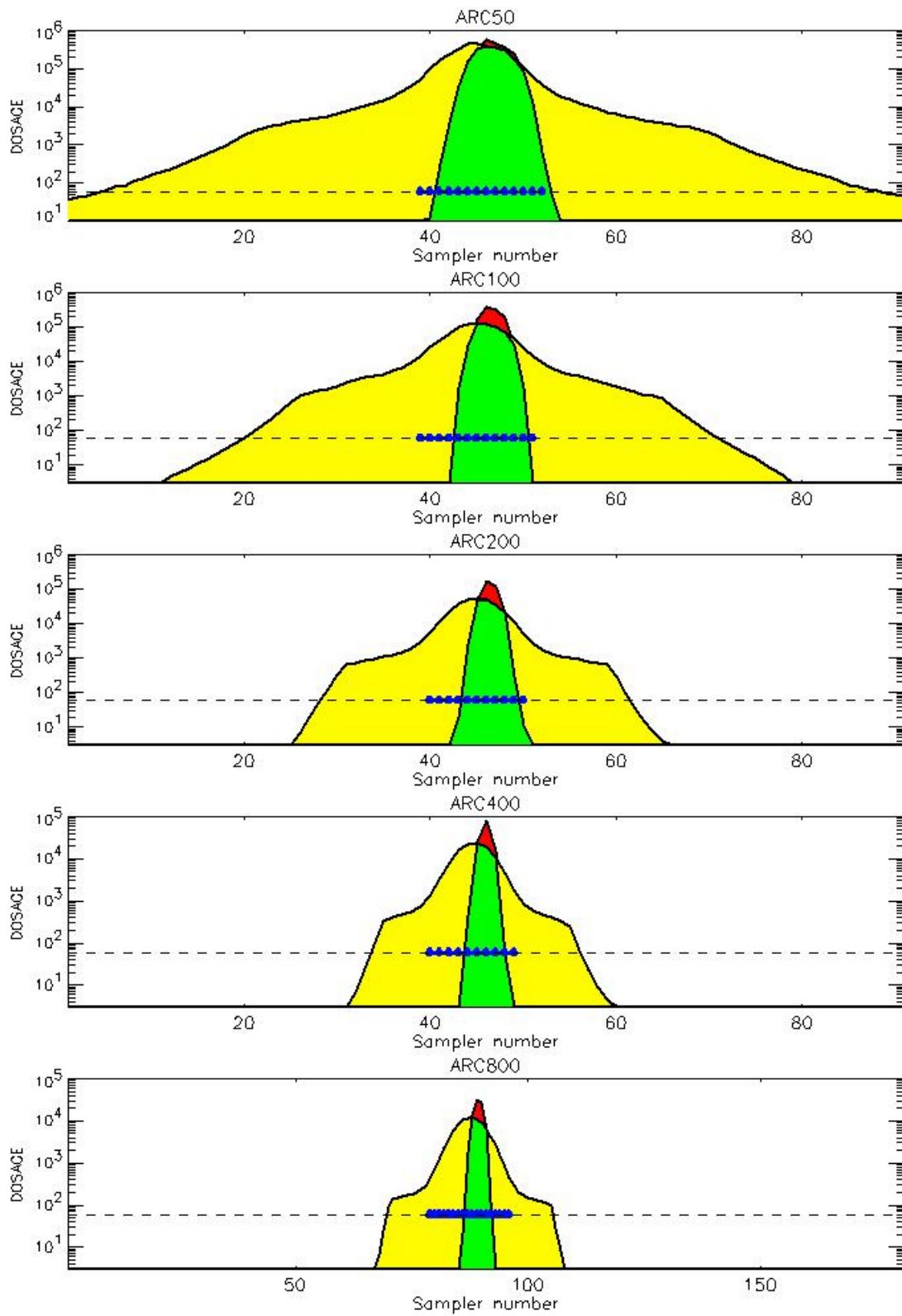


Figure 3-7. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Full Probability of Exceeding Threshold Value for Trial 58: Stability Class is 6

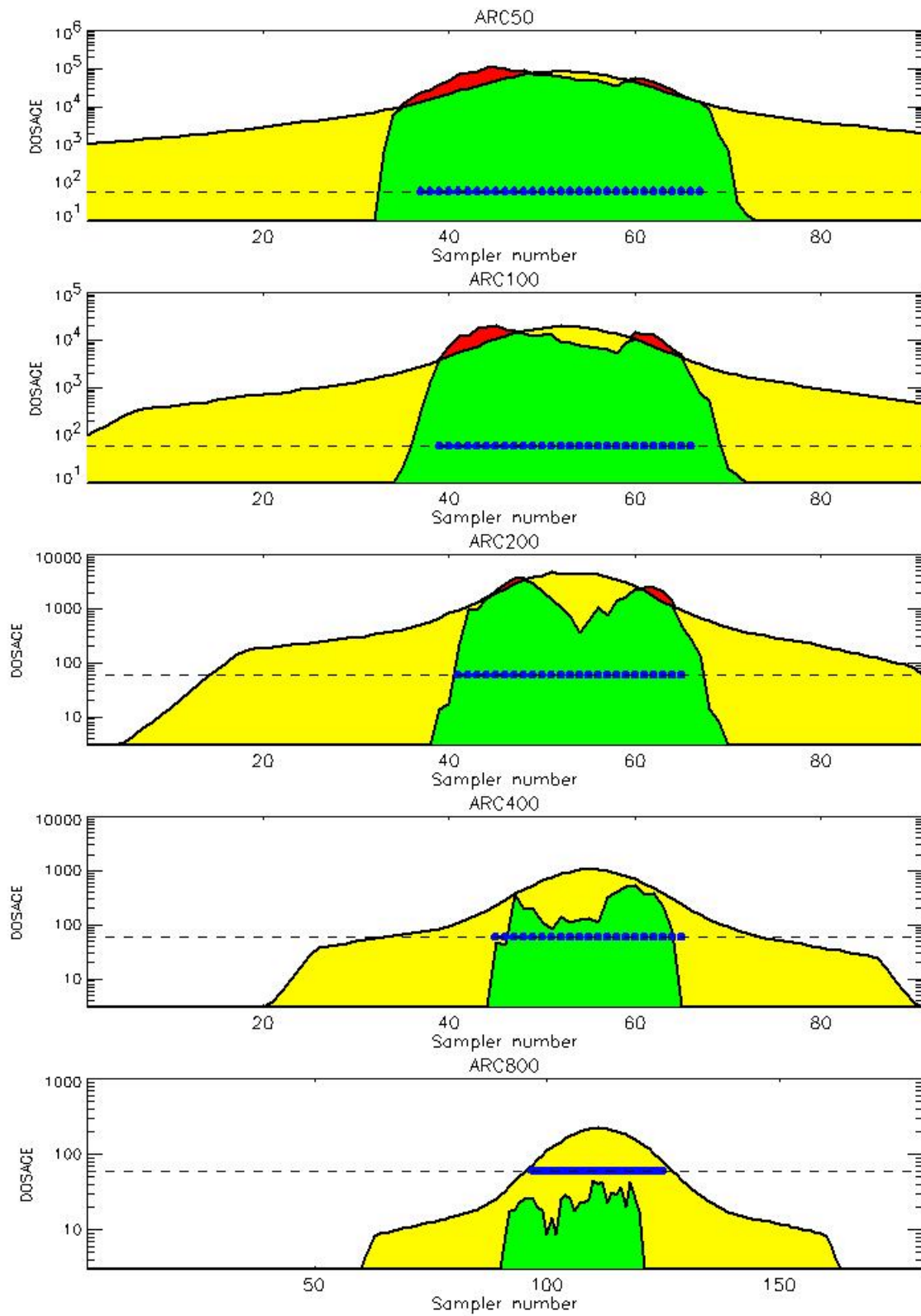


Figure 3-8. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Full Probability of Exceeding Threshold Value for Trial 16: Stability Class is 1

For the trials shown in Figures 3-7 and 3-8, the conditional probability prediction clearly leads to the wider plume at the specified threshold.⁹ (Compare the yellow region at 60 mg-sec/m³ to the extend of the blue diamonds.) An examination of the 51 figures in Appendix D shows that this is the general trend with the differences between the full and conditional probabilities being minimized for the longest-range arc, 800 meters.

The exclusion of intermittent spatial distributions, that is, the predicted zero dosage values at some locations, clearly increases the conditional probabilities relative to the full probabilities. For these data, the increases are maximized at shorter ranges and, for the investigated threshold value, lead to wider predicted plumes (relative to the full probability predictions). Figure 3-9 compares MOE values based on full and conditional probability predictions for the 50- and 800-meter arc. For the 800-meter arc, the conditional probability results are much more similar to the full probability results. The biggest differences between full and conditional probability MOE estimates are seen at the 50-meter arc. This is consistent with the expectation that, for longer ranges, dispersive effects tend to allow the plume to fill in pockets of zero dosage that can exist at shorter ranges.

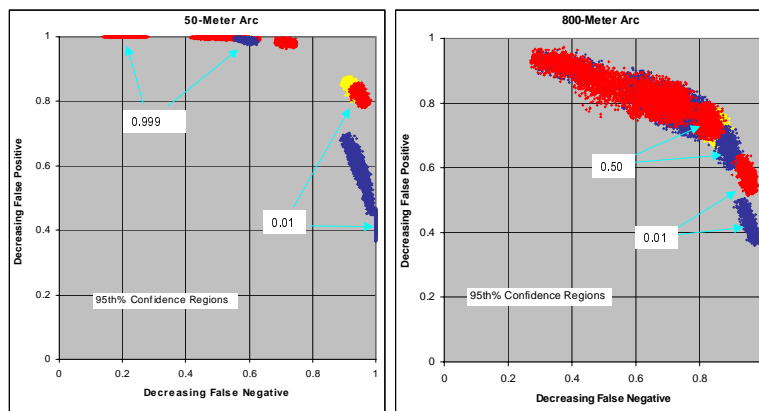


Figure 3-9. Direct Comparison of HPAC Full (Red Regions) and Conditional Probability (Blue Regions) Predictions (with Mean Value Prediction in Yellow): MOE Estimates (Based on 60 mg-sec/m³) for the 50- and 800-Meter Arc

⁹ Reference 3-2 reports an MOE estimate based on a threshold of 60 mg-sec/m³ for predictions of the *Prairie Grass* field trials by the Department of Energy's NARAC modeling system. For the same 51 trials as described above (e.g., in Figure 3-6), the NARAC-based MOE point estimate was (0.94, 0.62). The closest HPAC-based MOE result is the almost identical value of (0.95, 0.64) and is associated with the 0.01 full probability prediction. Therefore, in this sense, the HPAC 0.01 full probability predictions and the NARAC predictions of Reference 3-2 behave similarly, *on average*. However, we note, that as a function of SCG, there are substantial differences. For example, in the case with the largest differences (SCG = 4), the NARAC MOE point estimate is (1.00, 0.52) and the HPAC 0.01 full probability MOE point estimate is (0.81, 0.86).

2. MOEs Based on Summed Dosage

HPAC can provide the user with a dosage plume based on conditional probability predictions.¹⁰ The displayed values obtained in this way correspond to the dosages at which the conditional probability just exceeds the selected value (e.g., 0.85). Based on these dosages, MOE estimates can be generated. Figure 3-10 presents summed dosage-based MOE confidence regions for HPAC conditional probability predictions. The now familiar trend from 0.01 (lowest false negative and highest false positive) to 0.999 (highest false negative to lowest false positive) is shown.

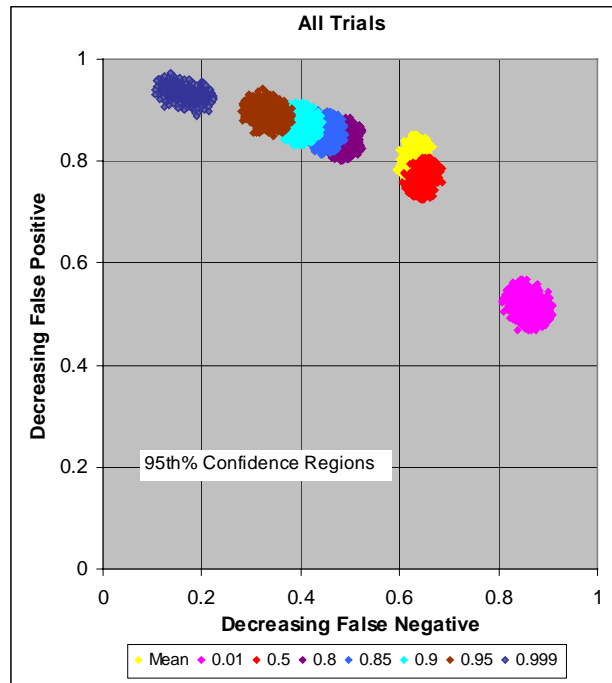


Figure 3-10. MOE Estimates (Based on Summed Dosage) for HPAC Conditional Probability Predictions: All Trials

For the summed dosage-based estimates, the 0.50 conditional probability MOE confidence region (red) overlaps with the mean value region (yellow) and these regions are closest to the (1,1). The MOE regions of Figure 3-10 suggest that the 0.50 conditional probability outputs performed best with respect to simply predicting the amount of material (i.e., the total dosage at the sampled height). That is, the MOE value associated with the 0.50 conditional probability predictions is closest, of the predictions examined, to the 45 degree diagonal.

¹⁰ For the full probability predictions, HPAC reports only probability values associated with some threshold value and *not* “dosage” plumes. Therefore, we did not compute MOE values based on summed dosages for the full probability predictions.

Appendix E provides 95th percent confidence regions for MOEs based on summed dosage as a function of arc range and SCG.

C. PROBABILISTIC OUTPUT RESULTS: BASED ON INTERPOLATED AREA

Figure 3-11 compares MOE confidence regions, based on a threshold of 60 mg-sec/m³, using three procedures.¹¹ The MOE estimates of the center plot are based on summing the distances along the sampler arcs as described in Chapter 2.A.1. The left plot relies on Delaunay triangulation to interpolate actual areas, and the right plot uses polar bilinear interpolation. Logarithmic interpolation of both the field trial observations and the predicted probabilities is accomplished using the Delaunay triangulation or the polar bilinear technique, as noted in Figure 3-11.¹²

The MOE estimates based on the three techniques are virtually indistinguishable from one another. Figure 3-11 shows that for the short-range, densely sampled, *Prairie Grass* field trial, the application of interpolation to obtain actual areas (e.g., in m²) leads to very few, perhaps not even noticeable, differences in MOE 95th percent confidence regions (at least for the case of full probability predictions).

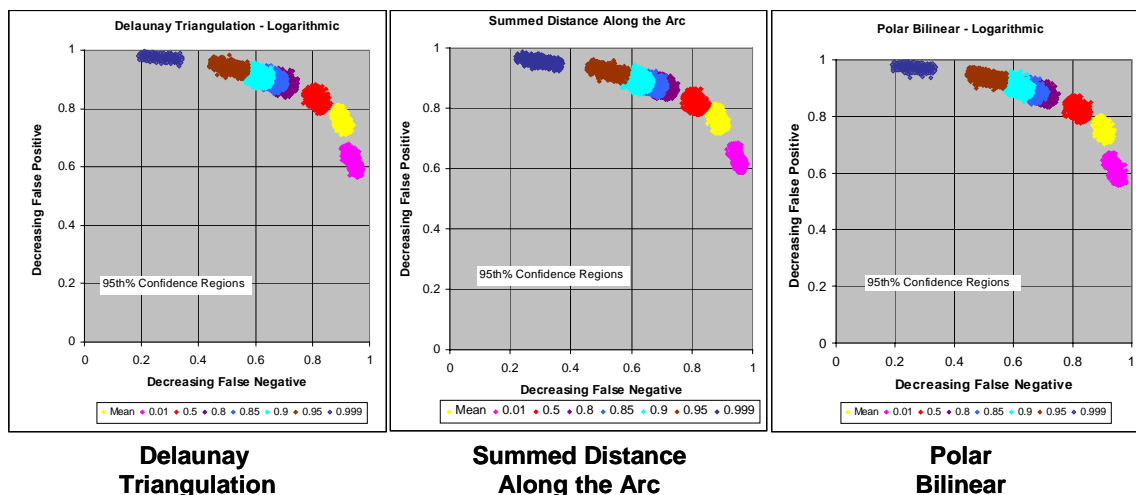


Figure 3-11. Comparison of “Summed Distance Along the Arc” and Interpolation Techniques: MOE Estimates (Based on 60 mg-sec/m³) for HPAC Full Probability Predictions and All Trials

¹¹ See Chapter 2, Section A.3 for additional discussion of interpolation techniques.

¹² Linear interpolation of the predicted probabilities is also examined in Appendix E and gives similar results.

Figures 3-12 and 3-13 compare results for HPAC conditional probability predictions based on interpolation techniques to those based on summed distances and summed dosages along the arc, respectively. Again, the MOE estimates are quite robust with respect to the interpolation technique used for these *short-range data*.

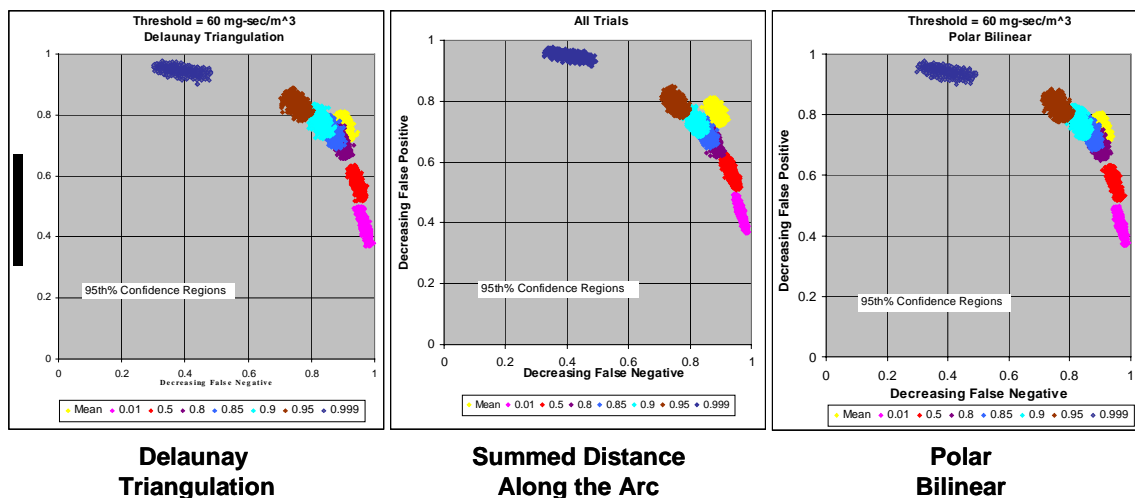


Figure 3-12. Comparison of “Summed Distance Along the Arc” and Interpolation Techniques: MOE Estimates (Based on 60 mg-sec/m³) for HPAC Conditional Probability Predictions and All Trials

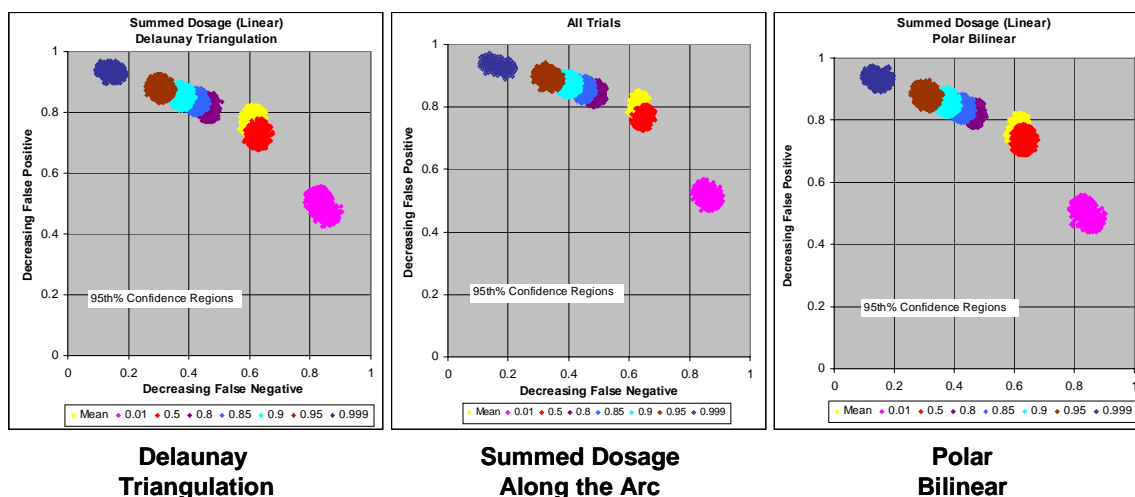


Figure 3-13. Comparison of “Summed Dosage Along the Arc” and Interpolation Techniques: MOE Estimates (Based on Summed Dosage – Linear Scale) for HPAC Conditional Probability Predictions and All Trials

Appendix E provides MOE confidence regions based on Delaunay triangulation and polar bilinear interpolation schemes and presents these results as a function of SCG. MOE estimates based on full and conditional probabilities are presented for both interpolation techniques.

For the above MOE plots that are based on “actual” areas obtained via interpolation, one might consider a spatially uniform population distribution. For this spatially uniform population, we define the “affected population” as those that reside within the bounds of the predicted or observed hazard area – here the area above the defined threshold (60 mg-sec/m^3). In such a case, we note that one can substitute the x-axis title “decreasing false negative” with something like “decreasing fraction of affected population inadvertently exposed,” and one might substitute the y-axis title with something like “decreasing fraction of the affected population unnecessarily evacuated.” With respect to user accreditation, area interpolation is of potential value because it scales the comparisons of predictions and observations (i.e., MOE evaluations) in terms of an affected population.¹³

D. MOE VALUES AND RISK TOLERANCE: USER-COLORED SPACE

This section provides notional evaluations of the computed MOEs via two user-defined FOMs, R-FOM and A-FOM (as described in Chapter 2.D). Figures 3-14 and 3-15 present overlays of MOE confidence regions on two user-coloring schemes of the 2D MOE space, R-FOM and A-FOM, respectively. The MOE estimates described in Figures 3-14 and 3-15 are based on HPAC full probability predictions.

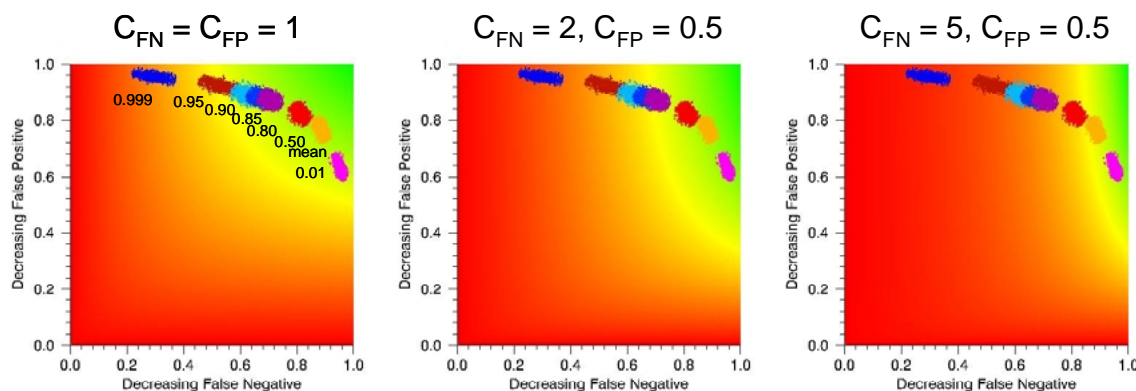


Figure 3-14. Overlay of MOE Estimates (Based on 60 mg-sec/m^3 Threshold) for HPAC Full Probability Predictions and All Trials: R-FOM

Based on the R-FOM user colorings described in Figure 3-14, the following evaluation is possible:

¹³ It is recognized that other sampler weighting schemes (e.g., perhaps based on the accuracy of a sampler’s measurements) might better serve other purposes (e.g., statistical model validation).

- For equal weighting of the false negative, false positive, and overlap fractions (i.e., $C_{FN} = C_{FP} = 1$), the full probability predictions in the range between 0.01 and 0.90 as well as the mean value predictions lead to acceptable (i.e., within the green user-colored space) model performance.
- For the somewhat conservative weighting that doubles the importance of the false negative fraction and halves the importance of false positive fraction ($C_{FN} = 2$, $C_{FP} = 0.5$) relative to the overlap fraction, the mean value predictions and the full probability predictions in the range 0.01 to 0.80 result in satisfactory model performance.
- Finally, for the very conservative $C_{FN} = 5$ and $C_{FP} = 0.5$, the 0.01 full probability prediction provides the only acceptable prediction (of those examined in this study). Of course, Figure 3-14 suggests that there is a range of probability values somewhere between 0.01 and 0.50 that should lead to acceptable performance for this user coloring.

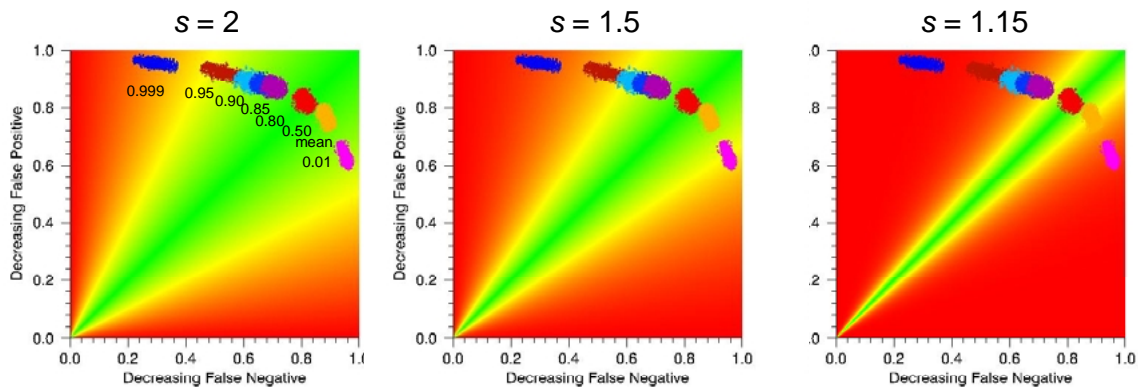


Figure 3-15. Overlay of MOE Estimates (Based on 60 mg-sec/m³ Threshold) for HPAC Full Probability Predictions and All Trials: A-FOM

Based on the A-FOM user colorings described in Figure 3-15, the following evaluations arise:

- For the requirement that the predicted and observed area sizes be, on average, within a factor of 2 (i.e., $s = 2$) of each other, several predictions appear potentially satisfactory (“in the green”). The mean value and full probability predictions in the range 0.01 to 0.95 appear satisfactory. The 0.999 full probability predictions do not appear acceptable. The 0.50 full probability predictions perform best by this criterion (of those examined).
- For the requirement that the predicted and observed area sizes be, on average, within 50 percent (i.e., $s = 1$) of each other, several predictions appear potentially satisfactory (e.g., mean value, 0.50, 0.80). Arguably, for the full probability results, the range 0.01 to 0.90 provides acceptable predictions.

The 0.50 full probability predictions perform best by this criterion (of those examined).

- Finally, for the strictest A-FOM requirement, $s = 1.15$, only the 0.50 full probability predictions lead to acceptable results (of those examined).

Figure 3-16 compares MOEs based on mean value, 0.50 full probability, and 0.50 conditional probability predictions for two user colorings. For the conservative R-FOM, the 0.50 conditional probability prediction is favored (at least slightly over the mean value prediction). The conservative nature of the conditional probability prediction, relative to the full probability prediction, is clearly evident in Figure 3-16. For the A-FOM coloring, the 0.50 full probability result performs the best (as also seen in Figure 3-15).

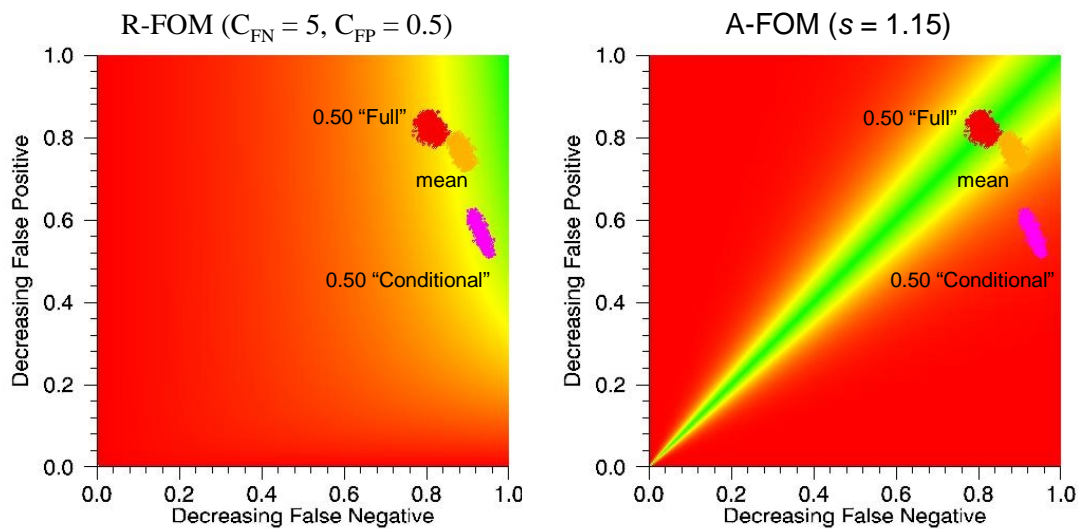


Figure 3-16. Overlay of MOE Estimates (Based on 60 mg-sec/m³ Threshold) for HPAC Full and Conditional Probability Predictions with All Trials: R-FOM ($C_{FN} = 5$, $C_{FP} = 0.5$) and A-FOM ($s = 1.15$)

Figure 3-17 considers the full probability predictions generated after polar bilinear interpolation. Since these MOEs were estimated based on actual false negative, false positive, and overlap *areas*, we can consider the potential affected population, at least for an assumed spatially uniform population. We define the affected population as that population that resides within either the area of the prediction (potential evacuees) or the area of the observation (potentially exposed). Further, we use a critical threshold of 60 mg-sec/m³. Given the above details, we can relabel the x- and y-axes, as illustrated in Figure 3-17.

The next section of this chapter further develops the notion of affected area and affected population.

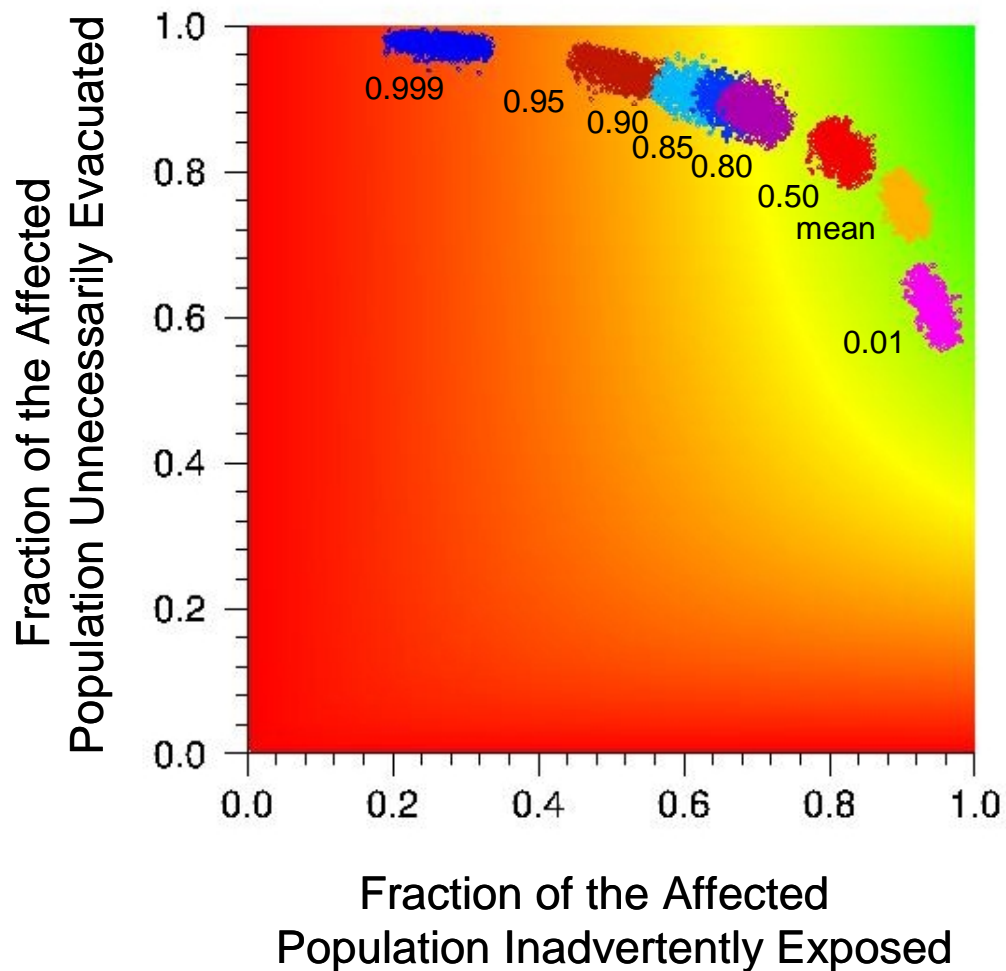


Figure 3-17. Overlay of MOE Estimates (Based on 60 mg-sec/m³ Threshold and Polar Bilinear Interpolation) for HPAC Full Probability Predictions and All Trials: R-FOM ($C_{FN} = 2$, $C_{FP} = 0.5$)

E. CONCEPT OF AVERAGE AFFECTED AREA

In this section we describe and develop the concept of average affected area. First, we note that, for a given prediction an estimate of the area – for example, above some threshold – can be developed. For some field trials or observations, it may be possible to develop a similar area. Certainly for the short-range *Prairie Grass* field trials, reasonable area estimates can be obtained via interpolation. Given these two areas, predicted and observed, we define the average affected area as follows:

$$\text{Average Affected Area} = \frac{A_{PR} + A_{OB}}{2} = \frac{A_{FN} + 2A_{OV} + A_{FP}}{2}. \quad (3-1)$$

Similarly, given a population distribution, one could compute average affected population. For example, for a spatially uniform population distribution, the average affected population and area are simply related by a constant. We denote this population as affected because they reside within either the predicted area of a hazard (prediction above some threshold) or because they reside within the observed hazard area (observation above the threshold). In some imagined scenario then, these affected people are subject to appropriate evacuation, unnecessary evacuation, or inadvertent exposure corresponding to A_{OV} , A_{FP} , and A_{FN} , respectively.

Next, we consider the two predictions and observations shown in Figure 3-18. In Scenario I, we see a relatively narrow plume, both predicted and observed. Perhaps, this situation is caused by stable atmospheric conditions during the release and subsequent transport. Scenario II illustrates a much wider and larger area of prediction and observation. This situation might arise from a larger amount of material released, relative to Scenario I, and/or unstable meteorological conditions. In addition, for this thought experiment, both scenarios lead to identical MOE values.

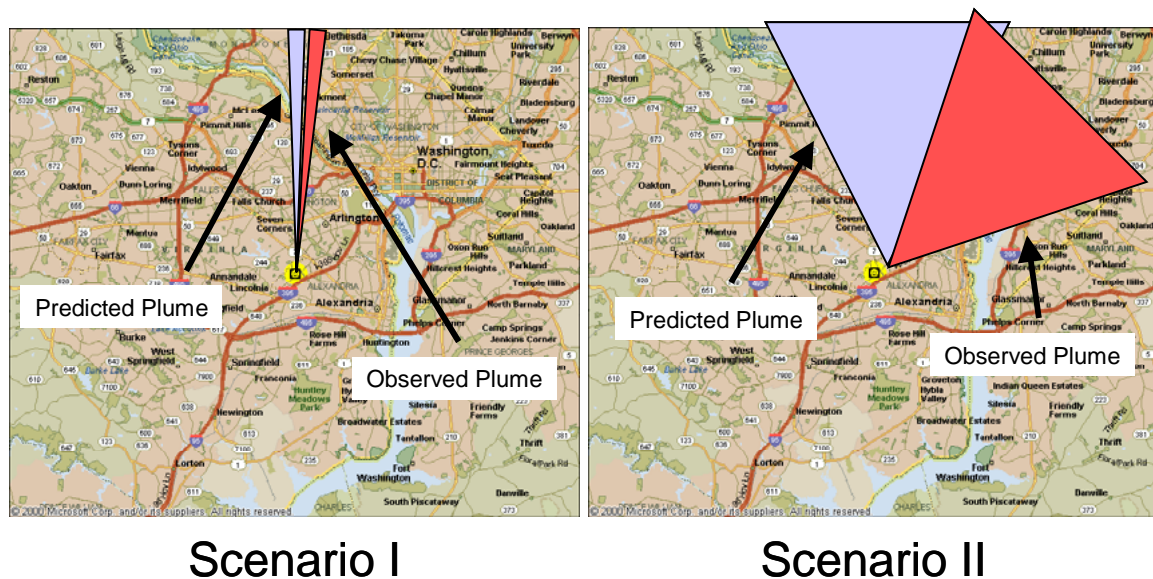


Figure 3-18. Two Notional Scenarios with “Identical” MOE Values

The implication is that the two scenarios lead to equal false negative and false positive *fractions*. From a validation, and perhaps accreditation, standpoint, it might be reasonable to consider the predictions of each scenario to perform equally well – they generate identical MOE values. However, an operational user, that is, someone who

makes a decision based on the model predictions, would be faced with very different situations. For example, a far larger evacuation effort would be needed for Scenario II.

For the HPAC mean value predictions of the 51 *Prairie Grass* field trials, we computed the average affected area. Figure 3-19 shows the distribution of these affected areas as a function of stability category grouping.¹⁴ The average affected areas are substantially larger for the trials conducted during under the more unstable conditions (SCG = 1,2,3) relative to those conducted during the more stable conditions (SCG = 4,5,6).¹⁵

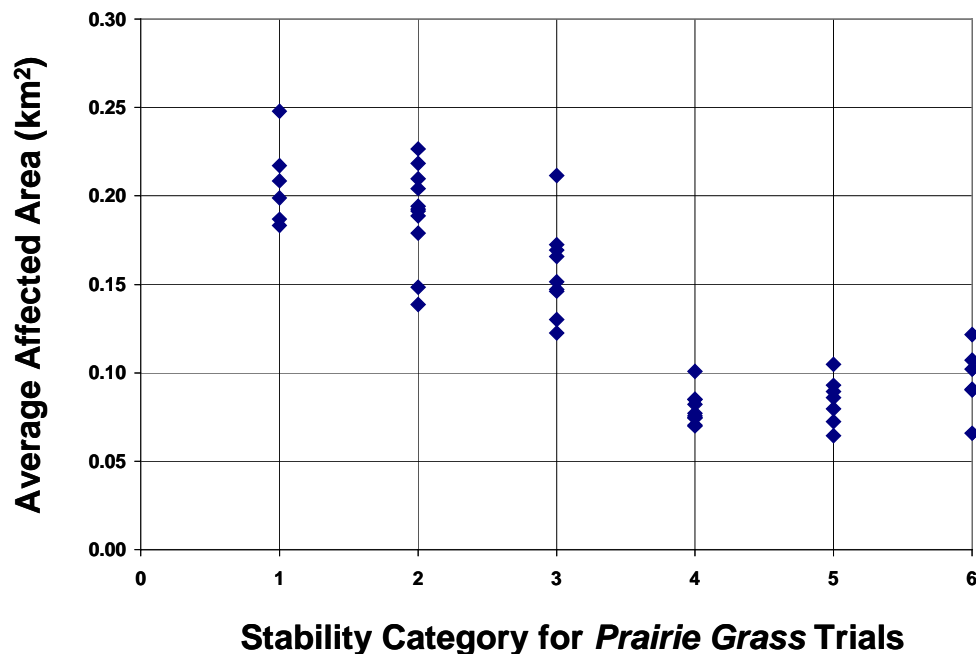


Figure 3-19. Average Affected Areas as a Function of Stability Category and Based on a 60 mg-sec/m³ Threshold and Delaunay Triangulation Interpolation

The average affected area values can be seen to separate the *Prairie Grass* predictions into two general categories, “stable” and “unstable.”

¹⁴ The two SCG 7 trials, Trials 13 and 14, are not included in this figure. These two trials had average affected areas of 0.155 and 0.192 km², respectively. For these two trials, the vertical dispersion was very limited, and it is believed that much of the SO₂ passed beneath the 1.5-meter above ground samplers along the 50-meter arc. Thus, one might consider these two trials as “outliers.”

¹⁵ Also, we found that for HPAC mean value predictions run with the uu(calm) toggle – a feature that determines the dispersion during the lightest wind conditions – set to 0.0 (our baseline for the comparisons of Reference 3-2) as opposed to the default value (0.25), the average affected areas associated with the stability categories 5 and 6 dropped somewhat. Therefore, the use of uu(calm) = 0.0 led to a consistent trend of decreasing average affected area with increasing stability category number.

Figure 3-20 provides perspective plots that show the 3-dimensional locations of the 49 *Prairie Grass* field trial predictions (excluding the two SCG = 7 trials). Each point's location is defined by its MOE value (x,y) and the average affected area (z). The predictions described by these points are based on HPAC mean value outputs. The MOE/affected area results shown in Figure 3-20 nicely separate into three stability category groupings: unstable/SCG = 1,2,3; neutral/SCG = 4; and stable/SCG = 5,6.

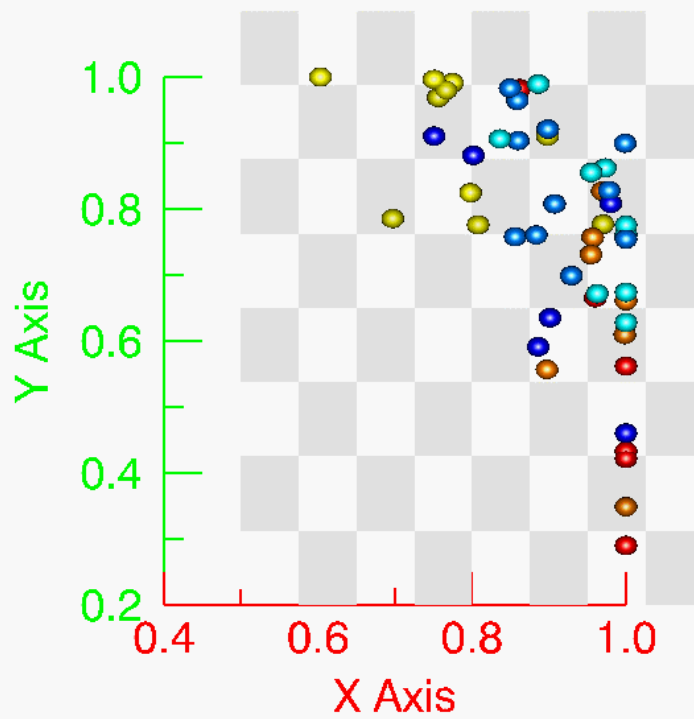
In Figure 3-20, each “ball” corresponds to the MOE/average affected area value for a particular *Prairie Grass* trial prediction. Stability class assignments are as follows: darkest blue = 1, blue = 2, lightest blue = 3, yellow = 4, orange = 5, and red = 6. The x-axis corresponds to “decreasing false negative,” that is, 1 minus the false negative fraction (A_{FN}/A_{OB}). The y-axis corresponds to “decreasing false positive,” that is, 1 minus the false positive fraction (A_{FP}/A_{PR}). Therefore, the perspective shown in Figure 3-20a, a look at the x-y plane, corresponds to the “normal” MOE view, albeit with MOE values based on individual trials vice averages with bootstrapped confidence regions. In the plots, the ranges of the x- and y-axes were limited to 0.4 to 1.0 and 0.2 to 1.0, respectively in order to expand the region of interest for viewing. The “checkerboard” sits on the x-y plane at $z = 0.12$ and will be discussed in a moment.

Figure 3-20b results from the rotation of the view shown in Figure 3-20a. In this view, the z-axis can be seen. The z-axis corresponds to the average affected area (in km^2).

Figure 3-21a, resulting from another rotation of Figure 3-20a, shows an end-on view of the checkerboard plane. From this perspective, it is clear that the average affected area values (the z-axis) separate the more stable trials (SCGs 4,5,6 = yellow, orange, red, respectively) from the more unstable trials (SCGs 1,2,3 = shades of blue). The unstable trials led to larger average affected areas – always greater than 0.12 km^2 . Alternatively, the stable trials led to z-values less than 0.12 km^2 .

Figure 3-21a (and also 3-20a and 3-20b) also indicates that the neutral trials (SCG 4 = yellow) separate (although not completely) from the stable trials (SCG 5,6 = orange and red) in the x-y (i.e., MOE) plane. Figure 3-21b also shows a somewhat different end-on view of the checkerboard plane. From this perspective view, it can be seen that the smallest average affected area associated with an “unstable” trial (marked in the figure) is about the same size as the largest average affected area associated with a “stable” trial. However, these two trials are well separated in MOE space (i.e., the x-y plane).

(a)



(b)

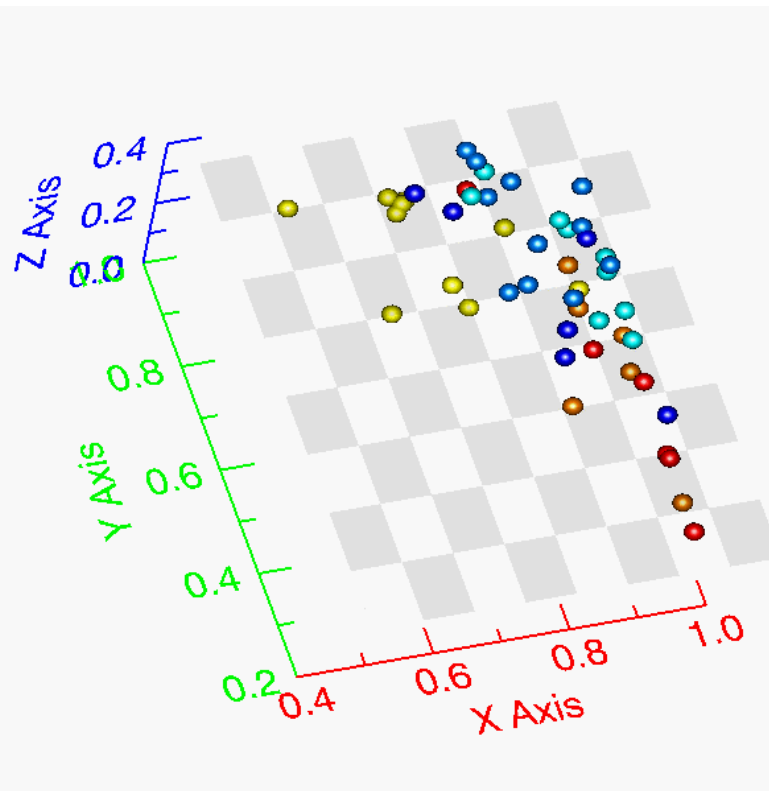
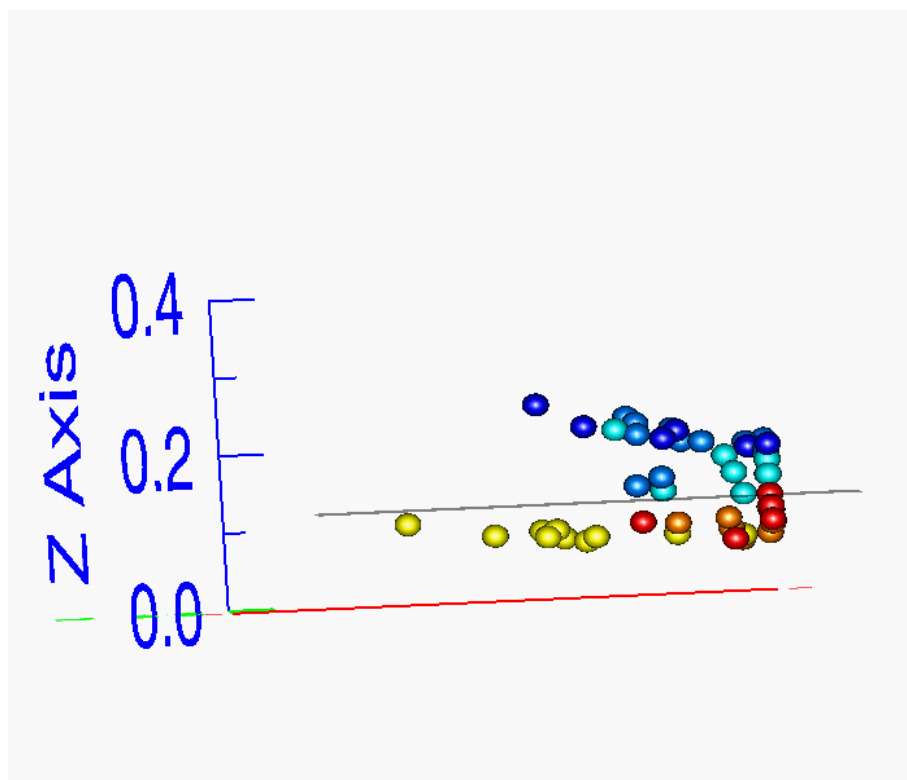


Figure 3-20. 3-Dimensional Locations (MOE and Average Affected Area) as a Function of Stability Category and Based on a **60 mg-sec/m³ Threshold** and Delaunay Triangulation Interpolation for **49 Individual Prairie Grass Field Trial Predictions**

(a)



(b)

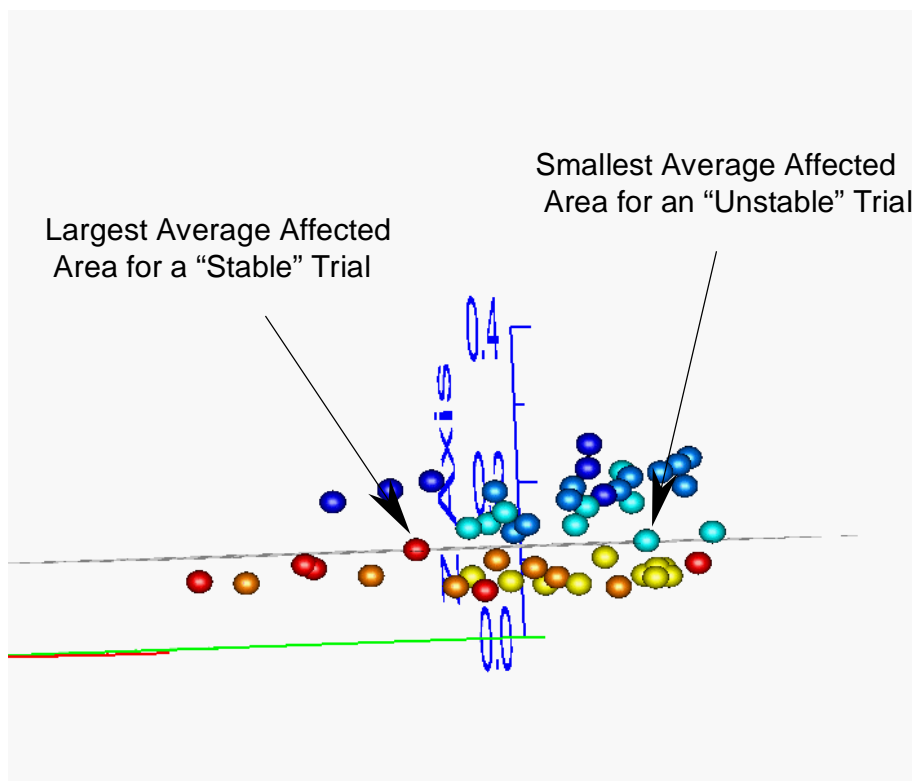


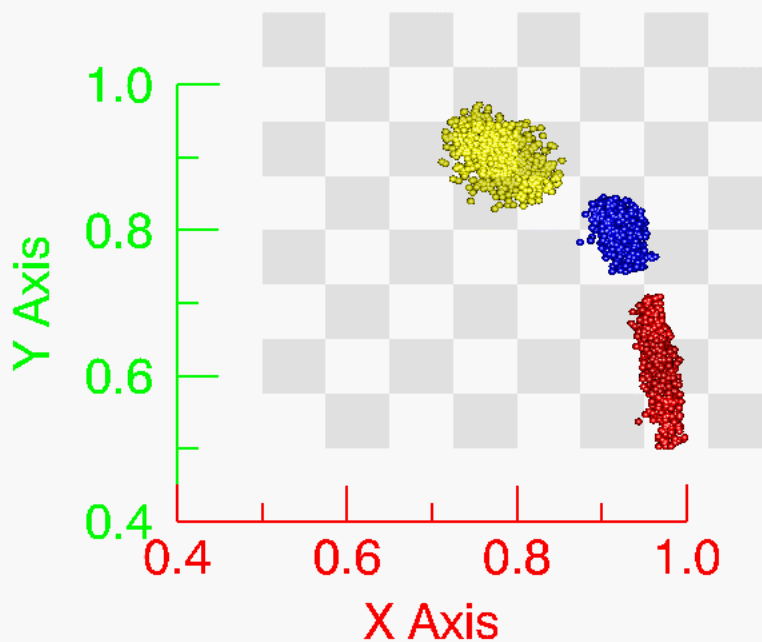
Figure 3-21. 3-Dimensional Locations (MOE and Average Affected Area) as a Function of Stability Category and Based on a **60 mg-sec/m³ Threshold** and Delaunay Triangulation Interpolation for **49 Individual *Prairie Grass* Field Trial Predictions: "End-On" Views**

The above apparently natural categorization of the *Prairie Grass* field trials – unstable, neutral, and stable – was used to group the data. Figure 3-22 shows 95th percent confidence regions for the MOE estimates for the three groups. The blue, yellow, and red confidence regions correspond to SCGs = 1,2,3; 4; and 5,6; respectively. Figure 3-22a shows the MOE perspective, that is, the x-y plane, and Figure 3-22b illustrates the separation along the average affected area axis (i.e., z-axis).

The following can be gleaned from Figure 3-22.

- The three groups – unstable, neutral, and stable – result in statistically different locations. (That is, the approximate 95th percent confidence regions are separate.)
- Predictions of the trials associated with the unstable conditions led to the MOE value closest to (1,1) and with a relatively large average affected area.
- Predictions of the trials associated with the neutral conditions led to the largest false negative fractions and a relatively small average affected area. For some users, the somewhat larger false negative fraction might be partially mitigated by the smaller affected area.
- Finally, predictions of the trials associated with the stable conditions led to a small false negative fraction, the largest false positive fraction, a relatively small average affected area, and overall the worst MOE value (“furthest from (1,1)”). For some users, the larger false positive fraction might be partially mitigated by the smaller affected area.

(a)



(b)

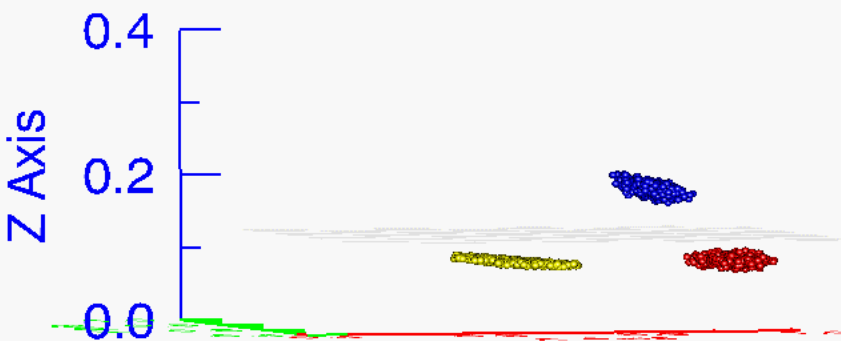


Figure 3-22. MOE (95th Percent Confidence) Regions and Average Affected Area as a Function of Stability Category and Based on a **60 mg-sec/m³ Threshold** and Delaunay Triangulation Interpolation for the *Prairie Grass* Field Trial Predictions

REFERENCES

- 3-1. Irwin, J. S. and Rosu, M-R., "Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models," *Proceedings of the 10th Joint Conference on the Applications of Air Pollution Meteorology*, American Meteorological Society, Boston, pp. 6-10, 1998.
- 3-2. Pasquill, F., *Atmospheric Diffusion, The Dispersion of Windborne Material from Industrial and other Sources*, Second Edition, Wiley, 1974.
- 3-3. Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T, and Larson, D., *User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models*, IDA Paper P-3554, January 2001.
- 3-4. Green, D. M., Swets, J. A., *Signal Detection Theory and Psychophysics*, Peninsula Publishing, Los Altos, CA, Reprint Edition, 1988 and Swets, J. A. editor, *Signal Detection and Recognition by Human Observers*, Peninsula Publishing, Los Altos, CA, Reprint Edition, 1988.

APPENDIX A

ACRONYMS

APPENDIX A

ACRONYMS

2D	Two-dimensional
A_{OB}	Area Associated With the Observations
A_{FP}	False Positive Area
A_{FN}	False Negative Area
A&M	Texas A&M University
A_{OV}	Area of Overlap
A_{PR}	Area Associated With the Prediction
ARAP	Aeronautical Research Associates of Princeton
ASCII	American Standard Code for Information Interchange
A_T	Total Area
C_{FN}	false negative coefficient
C_{FP}	false positive coefficient
cm	centimeters
DOD	Department of Defense
DOE	Department of Energy
DTRA	Defense Threat Reduction Agency
ETEX	European Tracer Experiment
FOM	Figure of Merit
GMU	George Mason University
HPAC	Hazard Prediction and Assessment Capability
hr	hour
IDA	Institute for Defense Analyses
IDL	Interactive Data Language
kg	kilogram
km	kilometer
L	Monin-Obukhov length scale
LLNL	Lawrence Livermore National Laboratory

μ	microns
m	meters
mg	milligrams
min	minutes
MIT	Massachusetts Institute of Technology
mm	millimeter
MOE	Measure of Effectiveness
m/s	meters per second
NARAC	National Atmospheric Release Advisory Center
NOAA	National Oceanic and Atmospheric Administration
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
ROC	Receiver Operating Characteristic
SCG	Stability Category Grouping
SCIPUFF	Second-Order Closure Integrated Puff
sec	seconds
SO ₂	Sulfur Dioxide
T_{avg}	Conditional Averaging
T&D	Transport and Dispersion
TWR	Tower
UTM	Universal Transverse Mercator
V&V	Verification and Validation
VV&A	Verification, Validation, and Accreditation
WMD	Weapons of Mass Destruction

APPENDIX B
SOME COMMENTS ON THE PREPARATION OF THE HPAC
PREDICTIONS USED IN THIS ANALYSIS

APPENDIX B

SOME COMMENTS ON THE PREPARATION OF THE HPAC PREDICTIONS USED IN THIS ANALYSIS

HPAC provides two convenient methods of producing data outputs that could be used for analysis: sampler output format and TAB (tabular output of the plotted data) output format. Given specific locations for the probes, the sampler output provides time series of mean concentration values (in [mass units]/m³), concentration fluctuation variance (in seconds), and integral time scale of the concentration fluctuation correlation (in seconds). Alternatively, given specific locations in the plume, TAB output format provides ASCII values of the quantities that are specified in the “Plot Choice” panel of the “Plot Control” window. For surface dosage plots, these quantities include “Mean Value (M),” “Probability (P[V>E]),” and “Exceed. Value (V[Pc>P])” options.¹

In a previous study [Ref. B-1], we summed up the mean concentration time series produced by the sampler output file format to calculate the mean dosages that were used in our analyses. In this paper, we deal with HPAC “probabilistic” outputs, and thus we decided to use TAB output file format for a number of reasons:

- For the TAB outputs (but not for the sampler file outputs), HPAC does the “work” of producing, that is, directly generates, the probabilistic values.
- We are validating HPAC outputs; thus we want to examine information that HPAC presents to users.

As a result of this switch from sampler output format to TAB output format we uncovered some programming bugs and inconsistencies between these outputs. These are described below.

¹ As described in Chapter 2, the “Probability (P[V>E])” option produces a “full probability” plume with values in the [0,1] range, while the “Exceed. Value (V[Pc>P])” option produces a dosage plume using the conditional probability of exceeding a specified probability in [0,1] range.

A. PRECISION OF TAB FILE INPUTS

HPAC does its calculation in single precision arithmetic. *Prairie Grass* field trials are represented in HPAC in the UTM coordinate system. Domain specifications are UTM zone 15, East-West Domain [534.20, 536.20] Easting, North-South Domain [4704.30, 4705.30] Northing. After running HPAC, we noticed that the TAB output format allows six decimal digits in specification of the probe (i.e., sampler) locations as shown in Figure B-1.

#	Northing	Easting	Data	
#=====	=====	=====	=====	
535.1500	4704.4270	535.150	4704.43	1.000000E-30
535.1500	4704.4287	535.150	4704.43	1.000000E-30
535.1501	4704.4305	535.150	4704.43	1.000000E-30
535.1503	4704.4322	535.150	4704.43	1.000000E-30
535.1505	4704.4340	535.151	4704.43	1.000000E-30
535.1508	4704.4357	535.151	4704.44	1.000000E-30

Figure B-1. Comparison of Specified Input Probe Locations for TAB Output Format (Left) With Those Actually Used in the Computation of Dosages (Right). Resolution of Input Locations of the Probes is 10 meters for the North-South Direction

The minimal north-south domain resolution is 10 meters. However, for a 50-meter arc, the inner sampler spacing is $r\theta = 50 \times 2 \times \frac{\pi}{180} \approx 1.745$ meters. Hence, a large number of samplers (especially at the 50-meter arc) are unresolved. To fix this problem, we shifted the origin of the HPAC simulations by 535 kilometers in the east-west direction and 4,704 kilometers in the north-south direction as shown in Figure B-2.

#	X (km)	Y (km)	Data
#=====	=====	=====	=====
	0.150000	0.427000	1.000000E-30
	0.150000	0.428700	1.000000E-30
	0.150100	0.430500	1.000000E-30
	0.150300	0.432200	1.000000E-30
	0.150500	0.434000	1.000000E-30
	0.150800	0.435700	1.000000E-30

Figure B-2. Shifting the Origin of HPAC Simulations by (535 km, 4704 km) Results in 1 mm Resolution of Input Locations for the Probes.

It would seem possible that a newer version of HPAC could convert geographic output domain to double precision arithmetic. As far as we know, the internal computational domain of HPAC is uncoupled from the geographic output domain; thus converting the geographic output domain to double precision would not affect the running time of HPAC significantly.²

B. TAB FILE “INFINITIES”

While using the TAB output format to obtain data, we discovered a software bug that produces unpredictable results and, under some circumstances,³ could completely stop HPAC calculations (“crash” the software). This bug is illustrated in Figure B-3.

```
#Project :pg_18_novd
#Path   :C:\HPAC\RUNS\Project2\
#Created :Fri Jan 26 10:34:21 2001
#Version :T:0.24-S:1.34
#Title(s):Prairie Grass Experiment #18
#       :Surface Dosage (1.50m)
#       :SO2 at 24-Jul-56 04:14Z (14.0 min)
#Units  :kg-sec/m3(x1.00E+06)
#Entries :2
#
#  X (km)    Y (km)      Data
#=====
# 0.297800   0.447801   1.000000E-30
# 0.297800   0.447800   INF
```

Figure B-3. Illustration of a Software Bug in HPAC’s TAB Output Format: Two Samplers are Separated by 1 mm, But the Second Dosage Value is INF

We have submitted an official HPAC “Bug Report” [Ref. B-2]. This software problem (bug) has been confirmed by the HPAC developers [Ref. B-3].

² Shifting the origin worked well in the *Prairie Grass* field trials because we did not include terrain in these calculations. Since, the *Prairie Grass* field trials took place over relatively flat terrain, and fields were mowed prior to trials, there is no need for either HPAC’s terrain or land usage databases. If, on the other hand, we had used either of the above databases, then shifting the origin would not work since these databases are coded to geographic coordinates.

³ By changing the units used for the plot display, one could obtain wildly oscillating values. When these probe locations are used with HPAC “probabilistic” outputs, they can cause the HPAC program to crash (pause without complete ion).

C. TAB VS. SAMPLER OUTPUT DIFFERENCES

Even after shifting the origin of the HPAC simulations by 535 kilometers in the east-west direction and 4,704 kilometers in the north-south direction, differences between the SAMPLER output format and TAB output format results still persisted as shown in Figure B-4. These differences are most noticeable at a 50-meter arc as depicted in Figure B-5. The dosage differences are at significant levels (approximately 50 percent of the maximum dosage obtained in the arc) and for some sampler locations are more than a factor of 2.

This problem was discussed with developers of HPAC. It was theorized that the main reason for this discrepancy is that SAMPLER file calculations are *exact*, while TAB calculations go through a series of *interpolations* with each step potentially adding error. Thus, technically speaking, it is possible that this is not a “coding” bug but an “algorithm” problem. This is potentially a problem. Here, there is the possibility of a user obtaining significantly different results for the same physical locations using different methods of reporting the data. This may represent a way for users to lose confidence in the model, and perhaps blame “unexplained” inconsistencies with a user's expectations on “bugs” in the model.

After submitting a bug report [Ref. B-4], we received a detailed reply from one of the developers of HPAC. It was postulated in that reply that the main reason for the observed discrepancies is a data resolution issue [Ref. B-5]. As stated in the reply:

“HPAC 3.2, by default, only uses data resolution equivalent to the screen pixel resolution. If the data has resolution beyond that of the screen pixel, the data is reduced⁴ up to the screen resolution. This is done to improve the speed at which plots are created and displayed. There are certain types of plots that require the use of the full resolution regardless of the screen resolution. For instance, if the user requests expected population estimates the plot will use the full data resolution available. The resolution is displayed in the lower right corner of the plot. If no resolution is displayed it means that the full data resolution as determined by the dispersion calculation is in use ... The data available for export, bitmap, contour or tabular, is the plot data. The HPAC display manager uses the available plot data to generate the export file.”

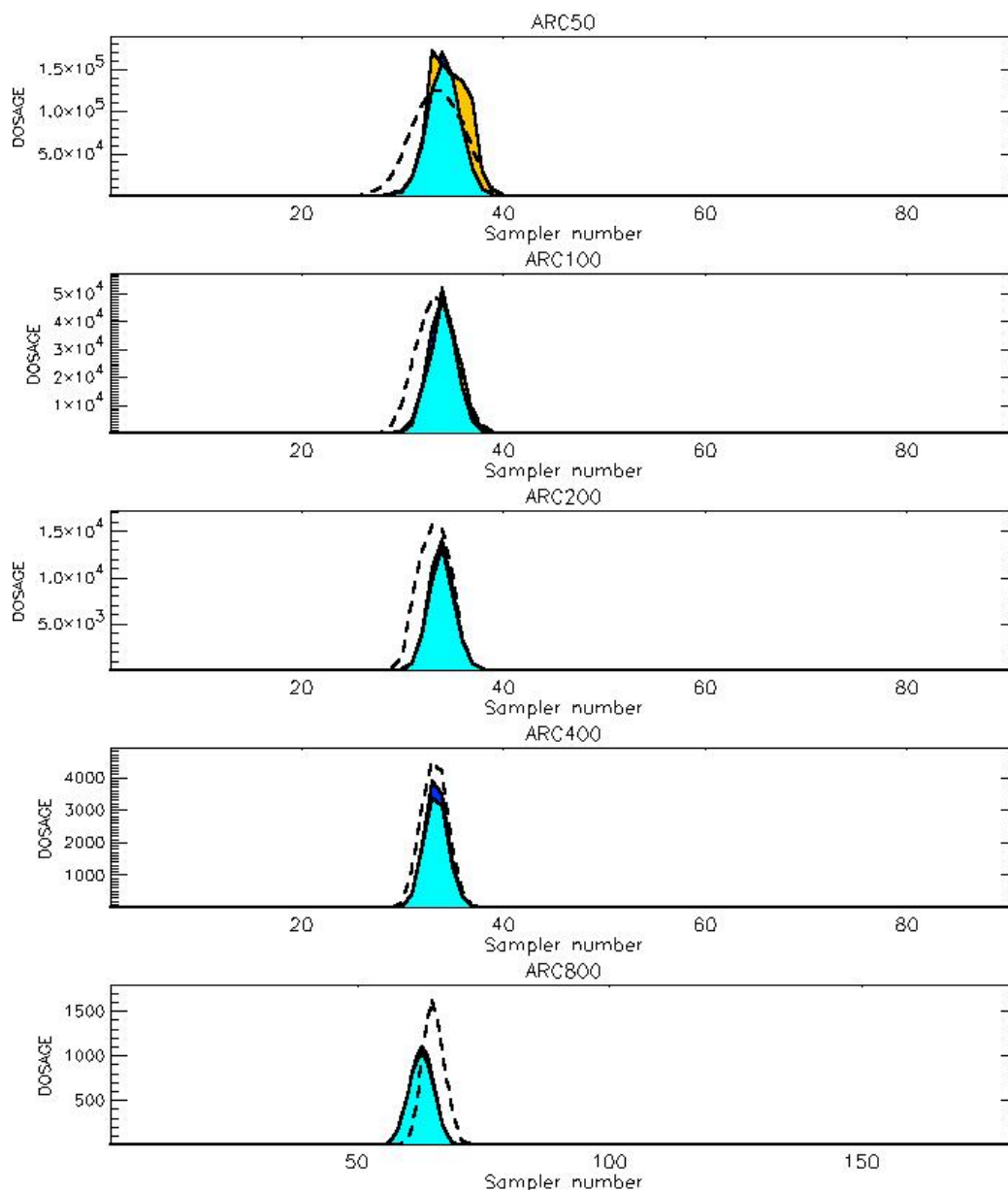
⁴ The actual word used in the reply was “dezoned,” but we believe that “reduced” is a clearer description of this procedure.

As stated in the reply, one can get full data resolution by requiring HPAC to calculate either “Area Estimate” or “Population Estimate.”⁵ The result of setting “Population Estimate” for the TAB output file format is depicted in Figure B-6. There is a noticeable improvement in terms of overall agreement between SMP and TAB file outputs. Nevertheless, there is still approximately 16 percent disagreement between the maximum values obtained using the TAB and SMP output file formats. Moreover, decreasing the time step by two orders of magnitude from 10 seconds to 0.1 seconds for HPAC’s calculations, in unison with the SMP output time interval, does not significantly narrow this difference. Thus, these differences are probably not related to the time stepping resolution.

D. 30 ORDER OF MAGNITUDE DROP-OFFS IN TAB PROBABILISTIC OUTPUTS

While using the “Exceed. Value ($V[P_c > P]$)” option that produces a dosage plume using the conditional probability of exceeding a specified probability in the [0,1] range, we noticed a very sharp drop-off in dosage values occurring over a short distance as shown in numerous plots in Appendix C and reproduced here in Figure B-7. Further examination of the TAB files shows that these drop-offs could span more than 30 orders of magnitude as shown in Figure B-8. While it is certainly feasible that such large dosage drop-offs could occur, it is highly unlikely that they would occur with such consistency. Since we believe that this problem could be related to the “Infinities” bug reported earlier in this appendix, we postponed reporting it to the developers of HPAC until resolution of the two bugs that were discussed earlier.

⁵ This requires checking one of the boxes inside “Maps” sub-menu of the “Plot” menu.



PG Prediction1 to Prediction2 Comparison

PG Trial File: `pr_grass_tracer_Experiment_55.txt`

PG Prediction File 1: `HPAC\novd_uucalm_0250\pg_55_novd.out`

PG Prediction File 2: `HPAC\tab\uu_calm_0250_mean\pg_55_novd.tab`

Figure B-4. Comparison of HPAC Predictions to *Prairie Grass* Trial 55 Using SAMPLER and TAB file Output Formats. The Dashed Line Represents Trial Dosages in mg-sec/m^3 , the Light Blue Region Represents Dosages That Agree for Both Sets of Predictions, the Light Brown Region Represents Dosages That are Higher for the TAB Output Format, and the Dark Blue Region Represents Dosages That are Higher for the SAMPLER Output Format. The Most Pronounced Differences in Dosages are at the 50-meter Arc.

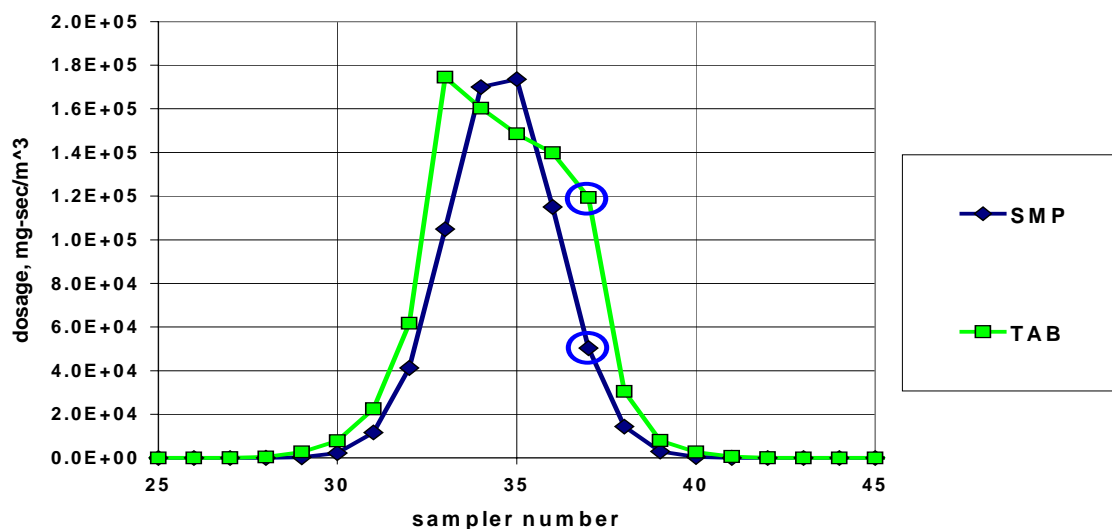


Figure B-5. Expanded Section of the 50-meter Arc of *Prairie Grass* Trial 55 Comparing HPAC's Predictions Using SAMPLER and TAB Output Formats. Blue Line Connects Dosage Values (in mg-sec/m³) Obtained Using SAMPLER Output Format and Green Line Connects Dosage Values Obtained Using TAB Output Format. Dosage Values at Circled Points Differ by More Than a Factor of 2.

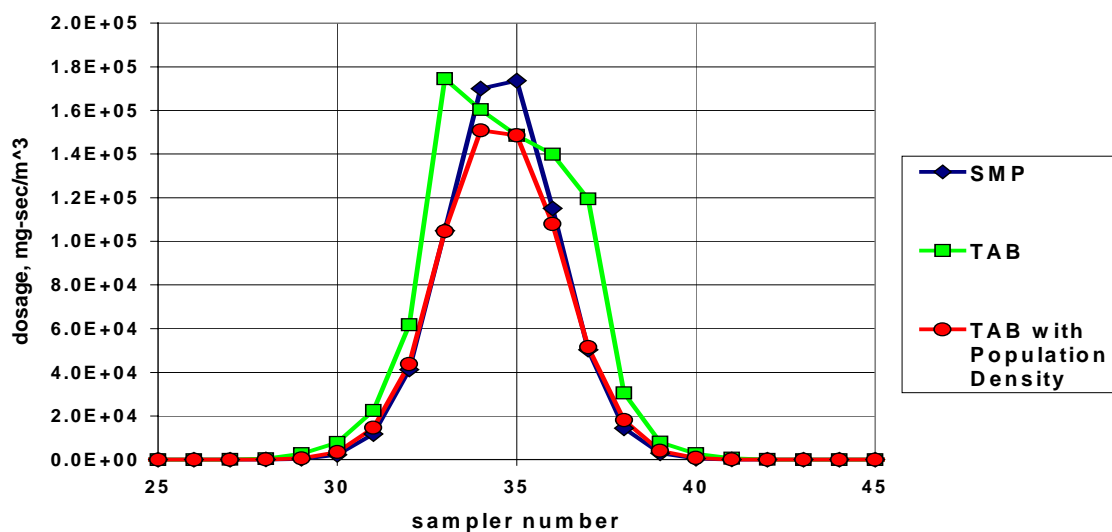


Figure B-6. Expanded Section of the 50-meter Arc of *Prairie Grass* Trial 55 Comparing HPAC's Predictions Using SAMPLER and TAB Output Formats. Blue Line Connects Dosage Values (in mg-sec/m³) Obtained Using SAMPLER Output Format, Green Line Connects Dosage Values Obtained Using TAB Output Format With Both Area Estimate and Population Estimate Flags Turned Off, and Red Line Connects Dosage Values Obtained Using Tab Output Format With Area Estimate Flag Turned On.

E. CONCLUDING REMARK

We conclude this appendix by showing that, as far as the user-oriented measures of effectiveness that were used in this paper are concerned, the dosage discrepancies that were noted earlier do not significantly alter our results. Figure B-9 compares HPAC predictions (by range) using the SAMPLER outputs format with the full UTM coordinate system specifications of the *Prairie Grass* field trials to the TAB output format using the shifted origin of the HPAC simulations (by 535 kilometers in the east-west direction and 4,704 kilometers in the north-south direction). The most significant differences occur at the 50-meter arc range, and even there, they are relatively minor.

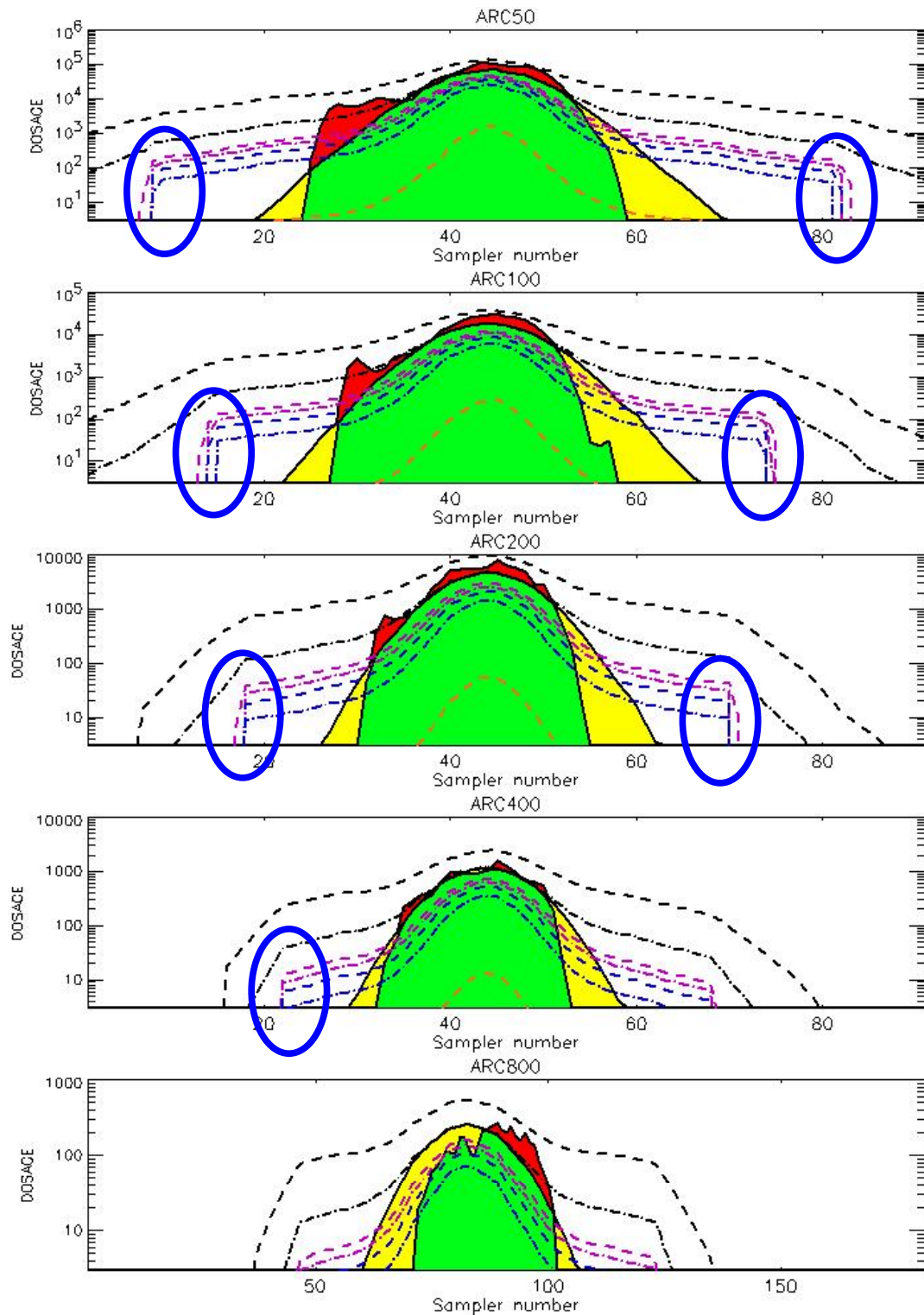


Figure B-7. HPAC Prediction to *Prairie Grass* Trial 5 on Logarithmic Scale Using “Conditional Probability” of Exceeding Specified Probability: Blue Ellipses Show Areas of “Suspicious” Dosage Drop-Offs.

```

#Project :pg_5_novd
#Path   :G:\HPAC-PG\NO_VD\
#Created :Thu Jan 04 17:15:59 2001
#Version :T:0.228-S:1.333
#Title(s):Prairie Grass Experiment #5
#       :Surface Dosage (1.50m)
#       :Pc(SO2) > 95.0% at 06-Jul-56 20:13Z (13.0 min)
#Units  :mg-sec/m3
#Entries :545
#
# X (km)    Y (km)    Data
#=====
0.150000    0.427000    1.000000E-30
0.150000    0.428700    1.000000E-30
0.150100    0.430500    1.000000E-30
0.150300    0.432200    1.000000E-30
0.150500    0.434000    1.000000E-30
0.150800    0.435700    1.000000E-30
0.151100    0.437400    1.000000E-30
0.151500    0.439100    20.0163
0.151900    0.440800    45.9776

```

Figure B-8. Extract of TAB Output File for HPAC Prediction of *Prairie Grass* Trial 5. The Two Samplers are Separated by Less Than 1.75 meters, While the Observed Dosage Drops-Off by More Than 30 Orders of Magnitude.

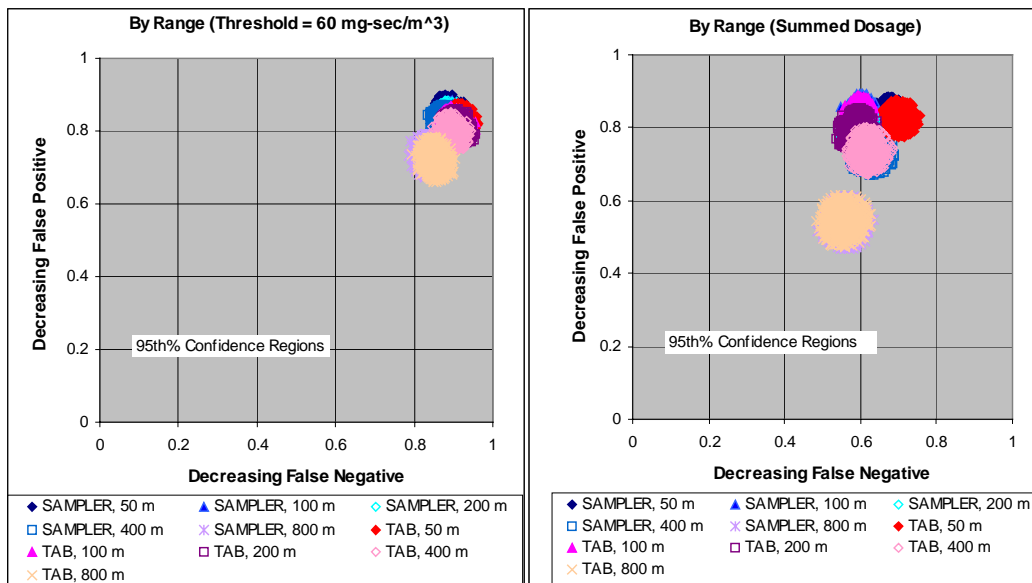


Figure B-9. User-Oriented MOE Comparison of HPAC Predictions Using SAMPLER and TAB Output Formats as a Function of Range: Based on Threshold = 60 mg-sec/m³ and Summed Dosage

REFERENCES

- B-1. Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T, and Larson, D., *User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models*, IDA Paper P-3554, January 2001.
- B-2. Platt, N., *Bug Report for TAB Export Option Resulting in INF Values*, e-mail for hpacbugs@saic.com, 2 February 2001.
- B-3. Stephens, T. C., SPCR 321-011: *Confirmation/Assignment: TAB Export Option Gives INV Values, Hangs HPAC*, e-mail to IDA and Titan Corporation 12 February 2001.
- B-4. Platt, N., *Bug Report for TAB and SMP OUTPUT Different Values*, e-mail for hpacbugs@saic.com, 1 March 2001.
- B-5. Parker, S. F., *Reply to Bug Report SPCR 321-013: TAB and SMP Output Different Values*, e-mail for SAIC and IDA, 2 April 2001.

APPENDIX C
HPAC PROBABILISTIC PREDICTION OUTPUTS COMPARED TO
PRAIRIE GRASS TRIALS

APPENDIX C

HPAC PROBABILISTIC PREDICTION OUTPUTS COMPARED TO *PRAIRIE GRASS* TRIALS

This appendix presents graphical comparison of HPAC probabilistic and mean value prediction outputs to *Prairie Grass* field trials [Ref. C-1].¹ Vertical plot units are dosage units of mg-sec/m³. Horizontal plot units are sampler numbers as presented in *the Prairie Grass* field trials with sampler number 1 oriented to the west, the middle sampler (45 or 90) oriented to the north, and the last sampler (91 or 181) oriented to the east of the SO₂ gas release source. Only data values greater than the cutoff threshold of 3 mg-sec/m³ (i.e., an average concentration of 0.005 mg/m³ over 600 seconds) are presented for both field trial data and HPAC probabilistic prediction outputs. This cutoff threshold value corresponds to a minimum value reported in *Prairie Grass* field trials.

Comparisons of HPAC probabilistic and mean value prediction outputs, as well as *Prairie Grass* field trial observations, are presented on both linear and logarithmic dosage scales. Each graphical comparison consists of five plots (one for each arc) with the top plot depicting the 50-meter arc, the second plot depicting the 100-meter arc, the third plot depicting the 200-meter arc, and so on. Odd-numbered pages contain figures on the linear dosage scale and even-numbered pages contain figures on the logarithmic dosage scale.

The solid lines correspond to the observed values and the HPAC mean value prediction. The meanings of the colors used in the plots are described in the below legend:

- Red — trial dosages that are higher than the mean value prediction
- Green — trial dosages that overlap with the mean value prediction
- Yellow — mean value predicted dosages that are higher than the trial observations
- White — trial sampler observation is missing.

¹ These predictions were done without the inclusion of an SO₂ surface deposition mechanism.

There are seven dashed lines that correspond to seven conditional probability prediction outputs generated by HPAC. These outputs are denoted “Exceed. Value ($V[P_c > P]$)” on the HPAC plotting screen in the *Plot Choice* dialog box. The dashed lines correspond to conditional probability values, from innermost line to outermost line, of 0.999, 0.95, 0.90, 0.85, 0.80, 0.50, and 0.01.

Samplers with missing values, or spurious maximum that were fixed by moving decimal point in the analyses, are noted in the figure captions [Ref. C-2]. Additional information is contained in Reference C-3.

Irwin stability categories that were used in our analyses are denoted in the figure captions. These stability category assignments are based on Reference C-4.

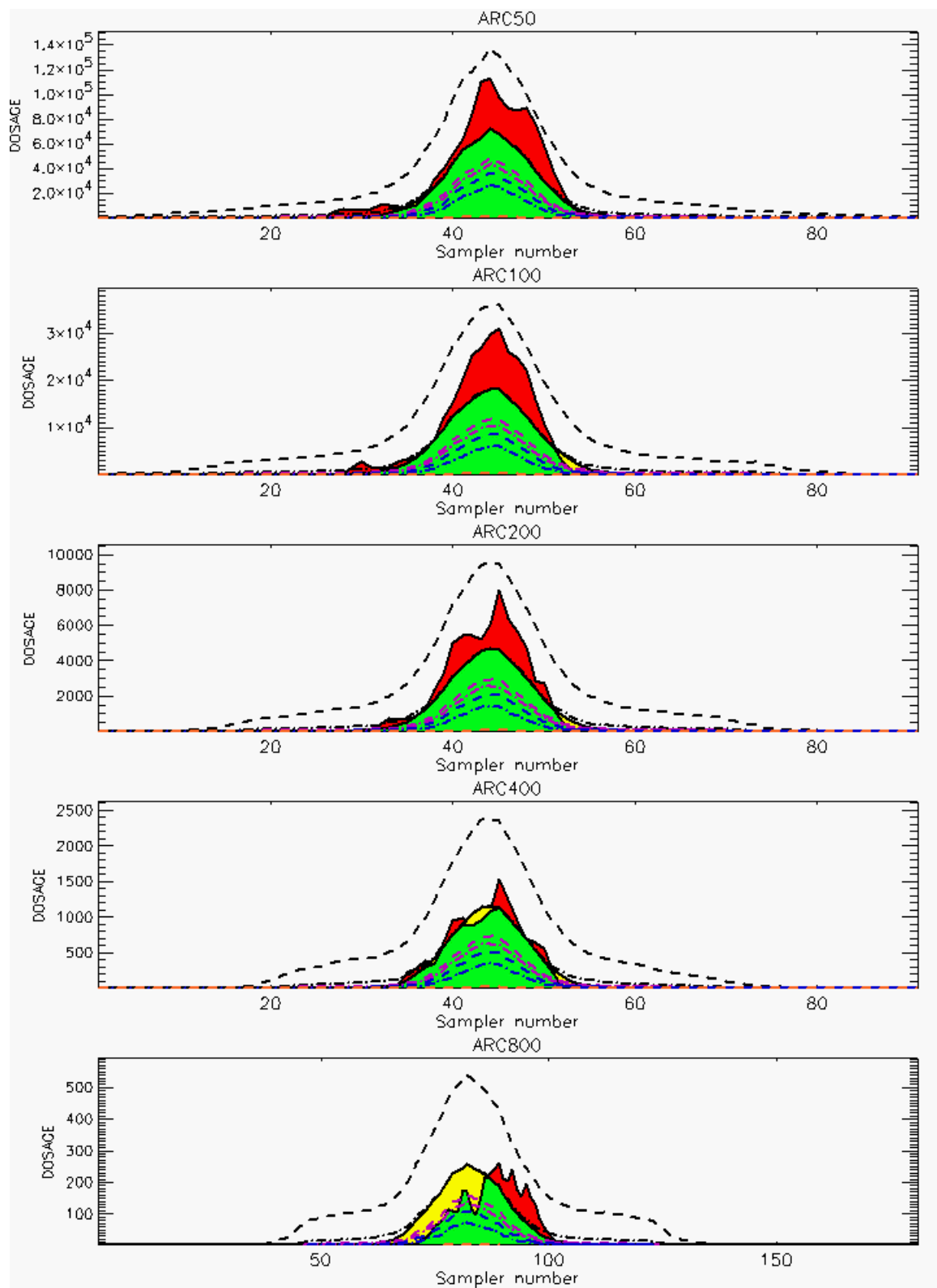


Figure C-1a. HPAC Probabilistic Prediction Outputs for Trial 5 on Linear Scale: Stability Class is 2

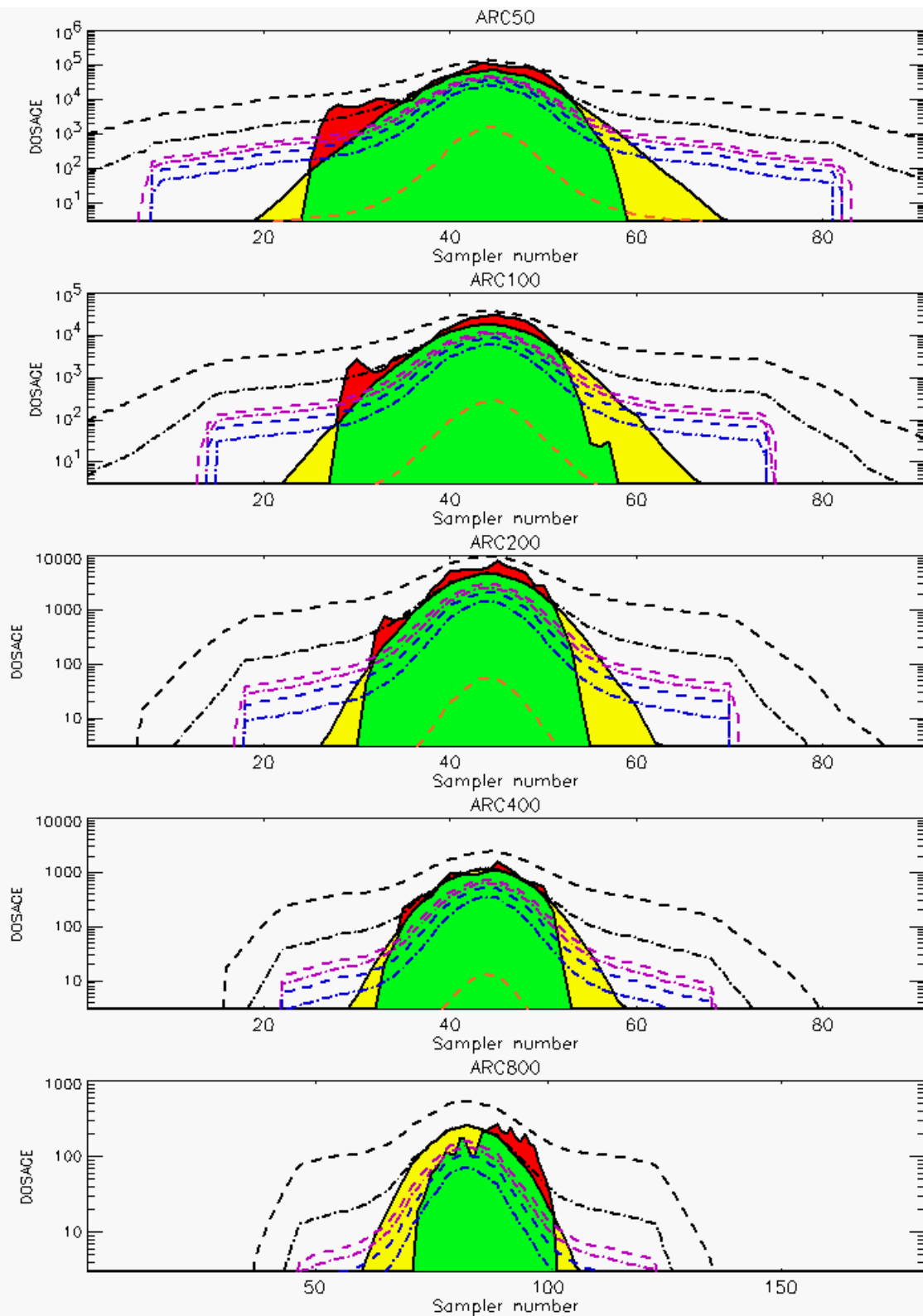


Figure C-1b. HPAC Probabilistic Prediction Outputs for Trial 5 on Logarithmic Scale: Stability Class is 2

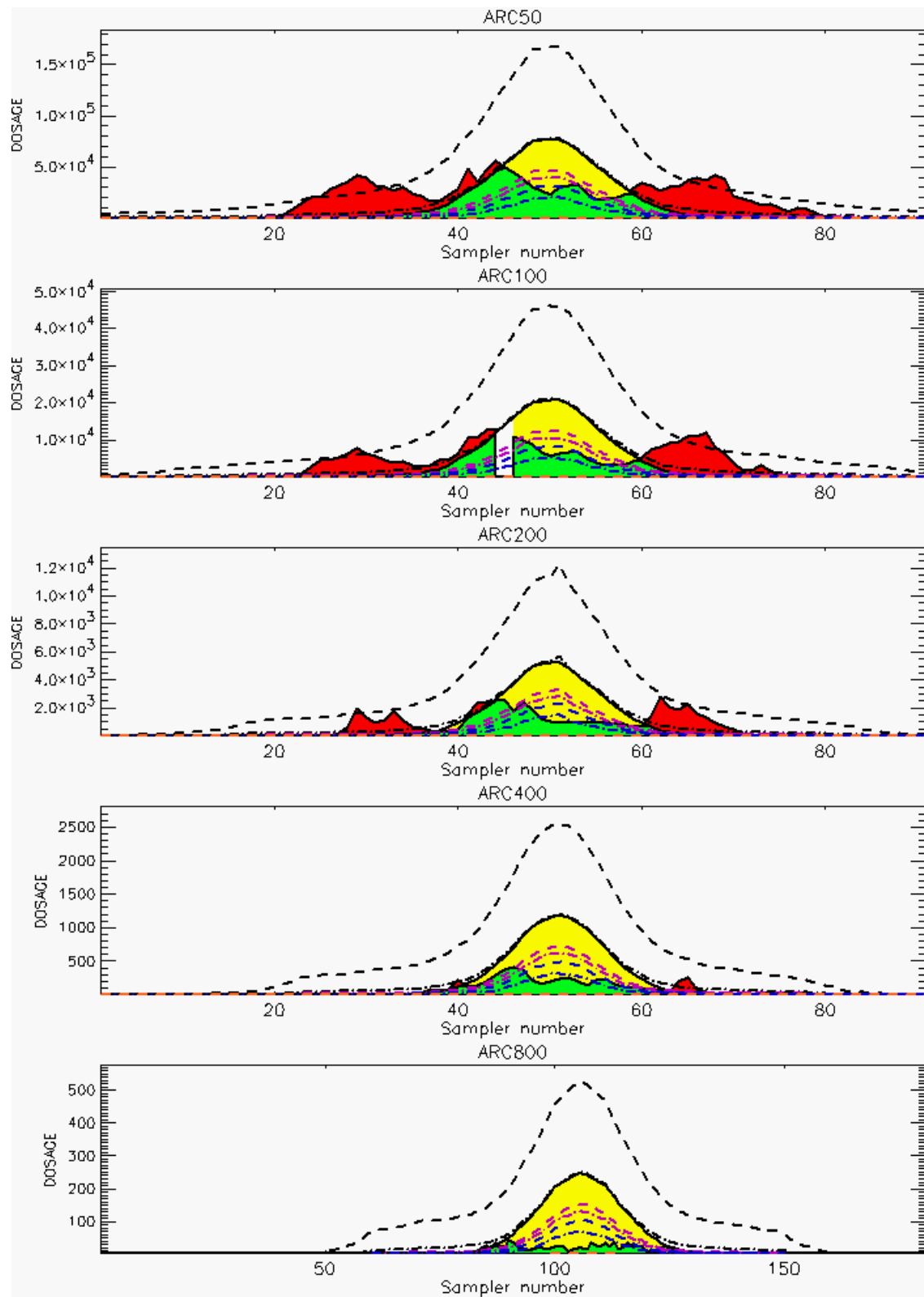


Figure C-2a. HPAC Probabilistic Prediction Outputs for Trial 7 on Linear Scale: Stability Class is 1 (Value for Sampler 45 of 100-Meter Arc is Missing)

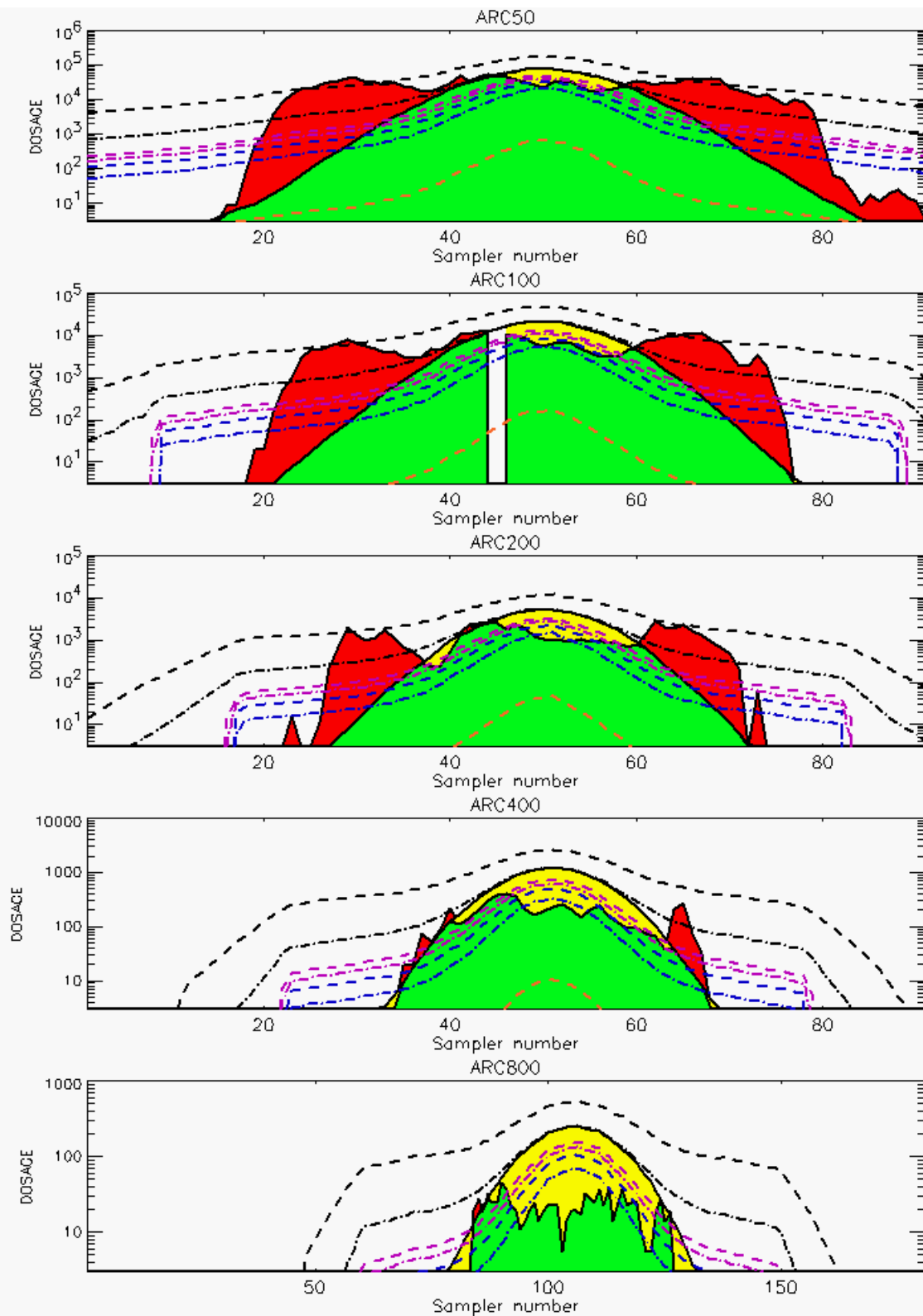


Figure C-2b. HPAC Probabilistic Prediction Outputs for Trial 7 on Logarithmic Scale: Stability Class is 1 (Value for Sampler 45 of 100-Meter Arc is Missing)

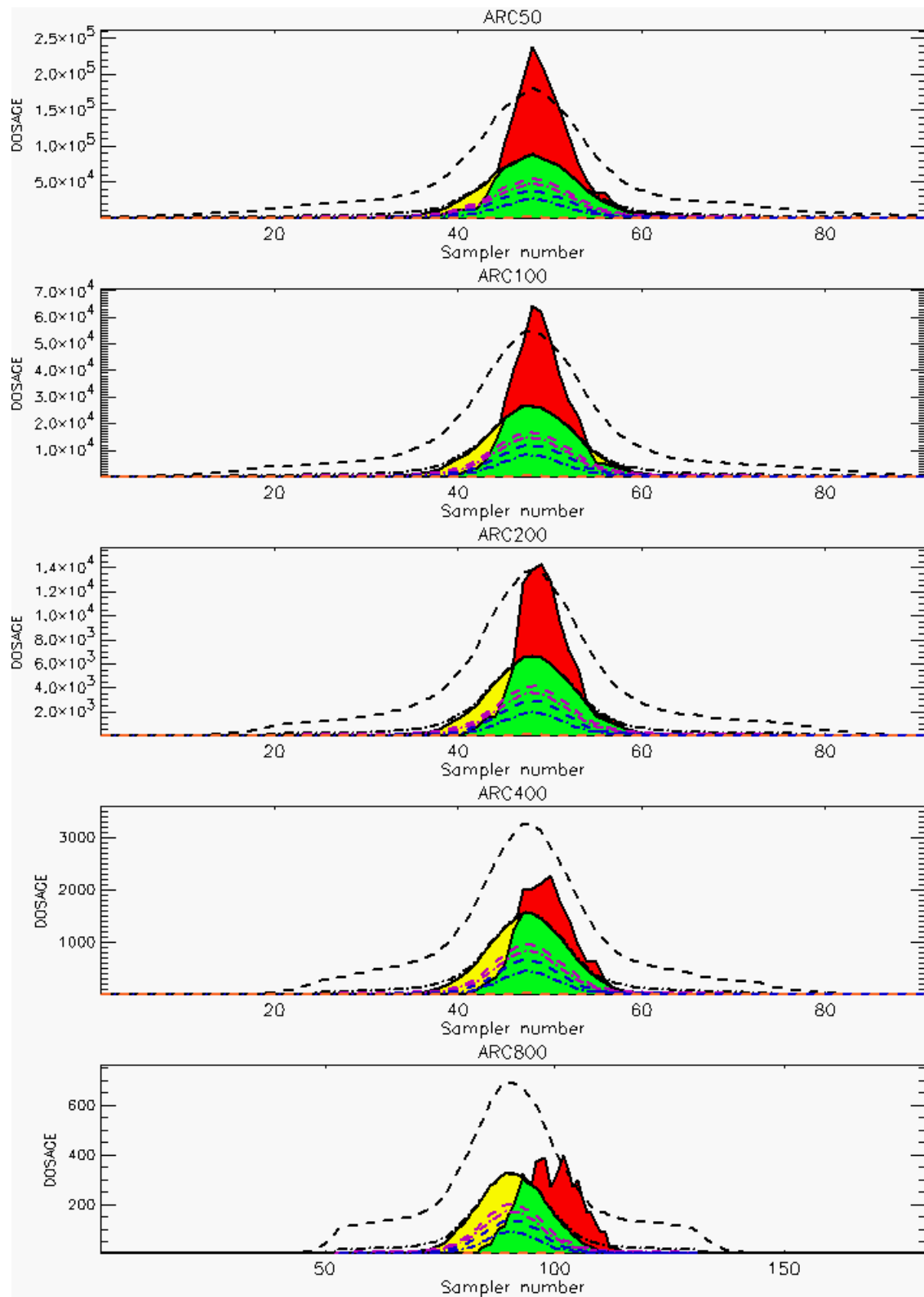


Figure C-3a. HPAC Probabilistic Prediction Outputs for Trial 8 on Linear Scale: Stability Class is 2

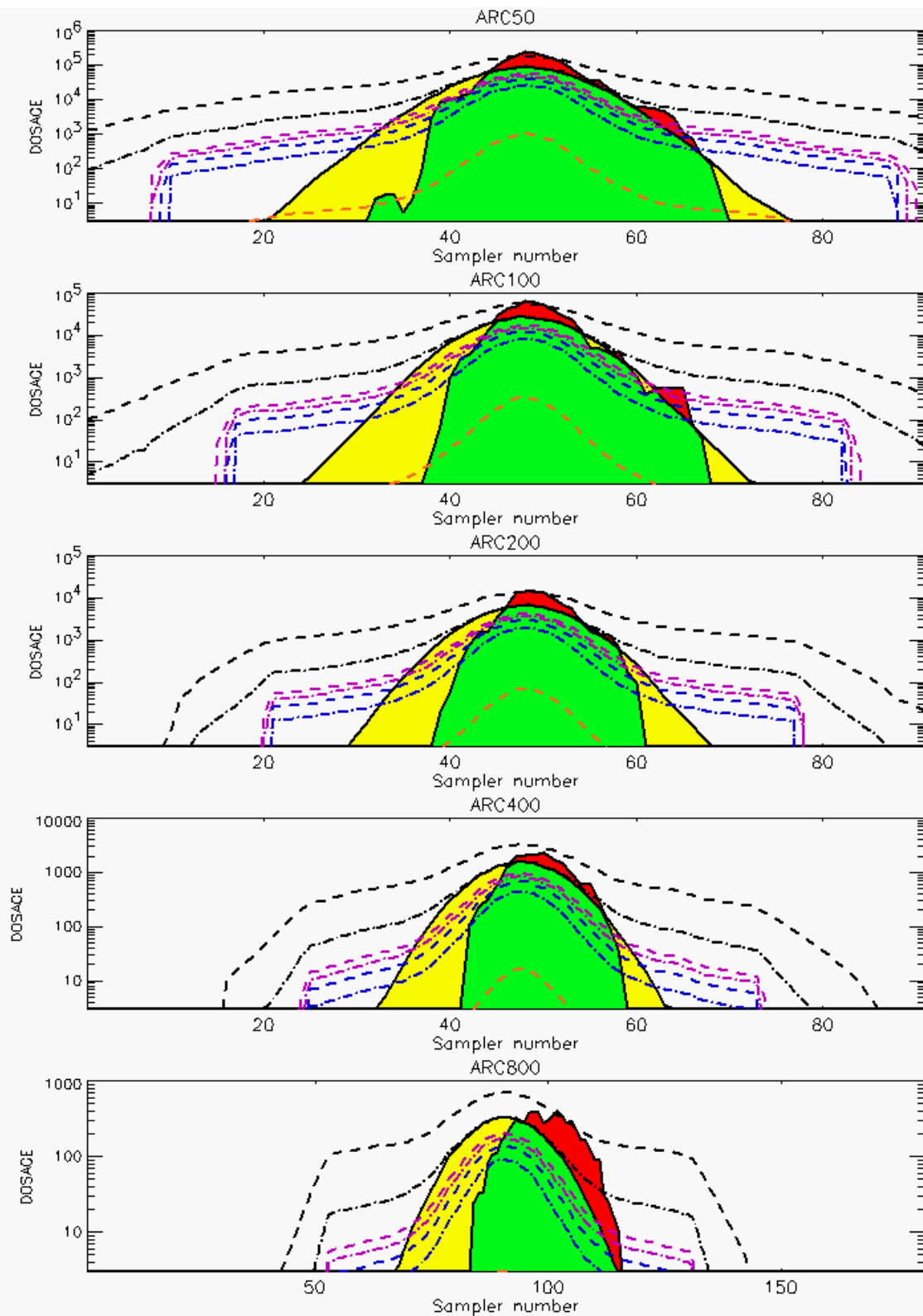


Figure C-3b. HPAC Probabilistic Prediction Outputs for Trial 8 on Logarithmic Scale: Stability Class is 2

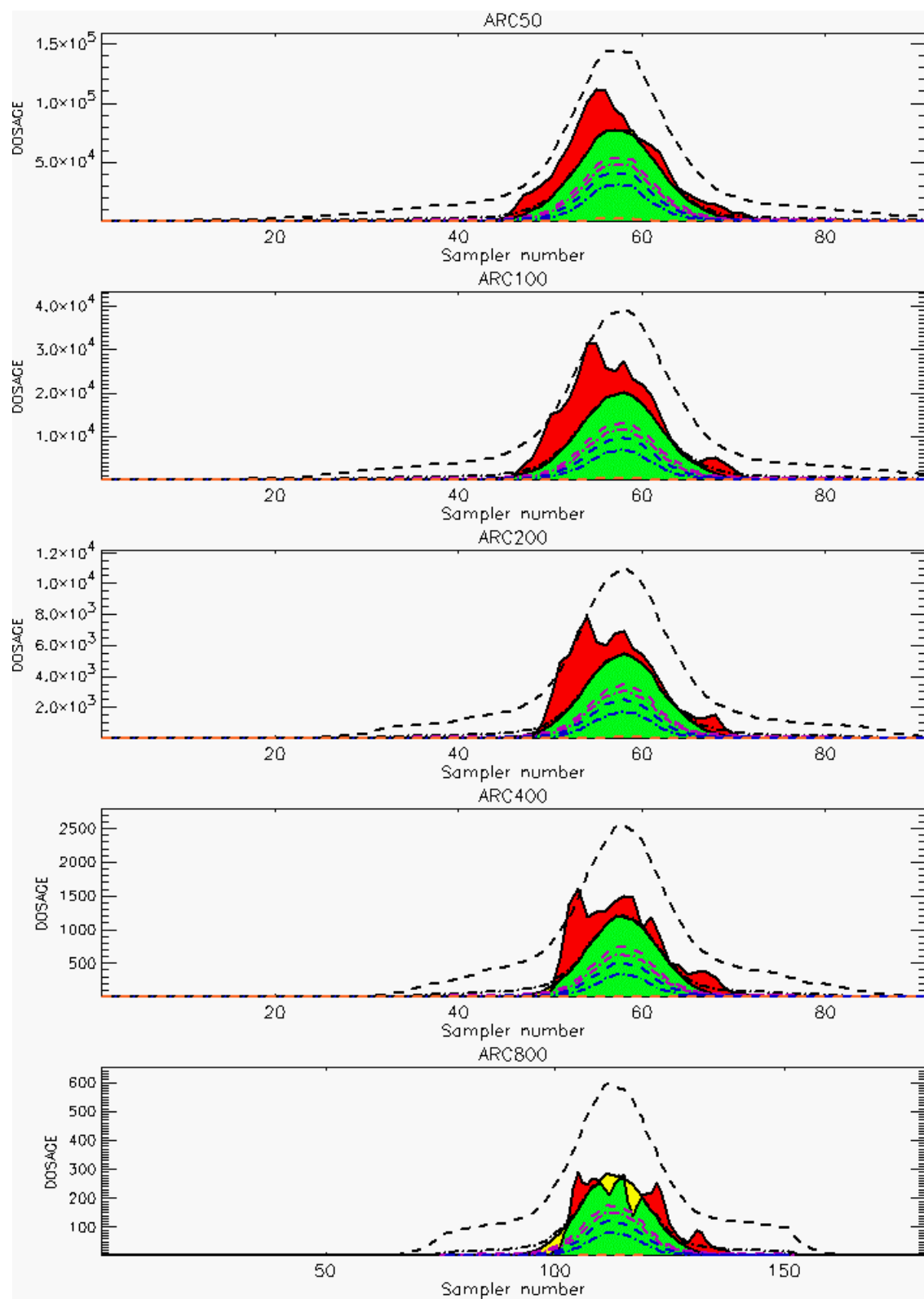


Figure C-4a. HPAC Probabilistic Prediction Outputs for Trial 9 on Linear Scale: Stability Class is 2

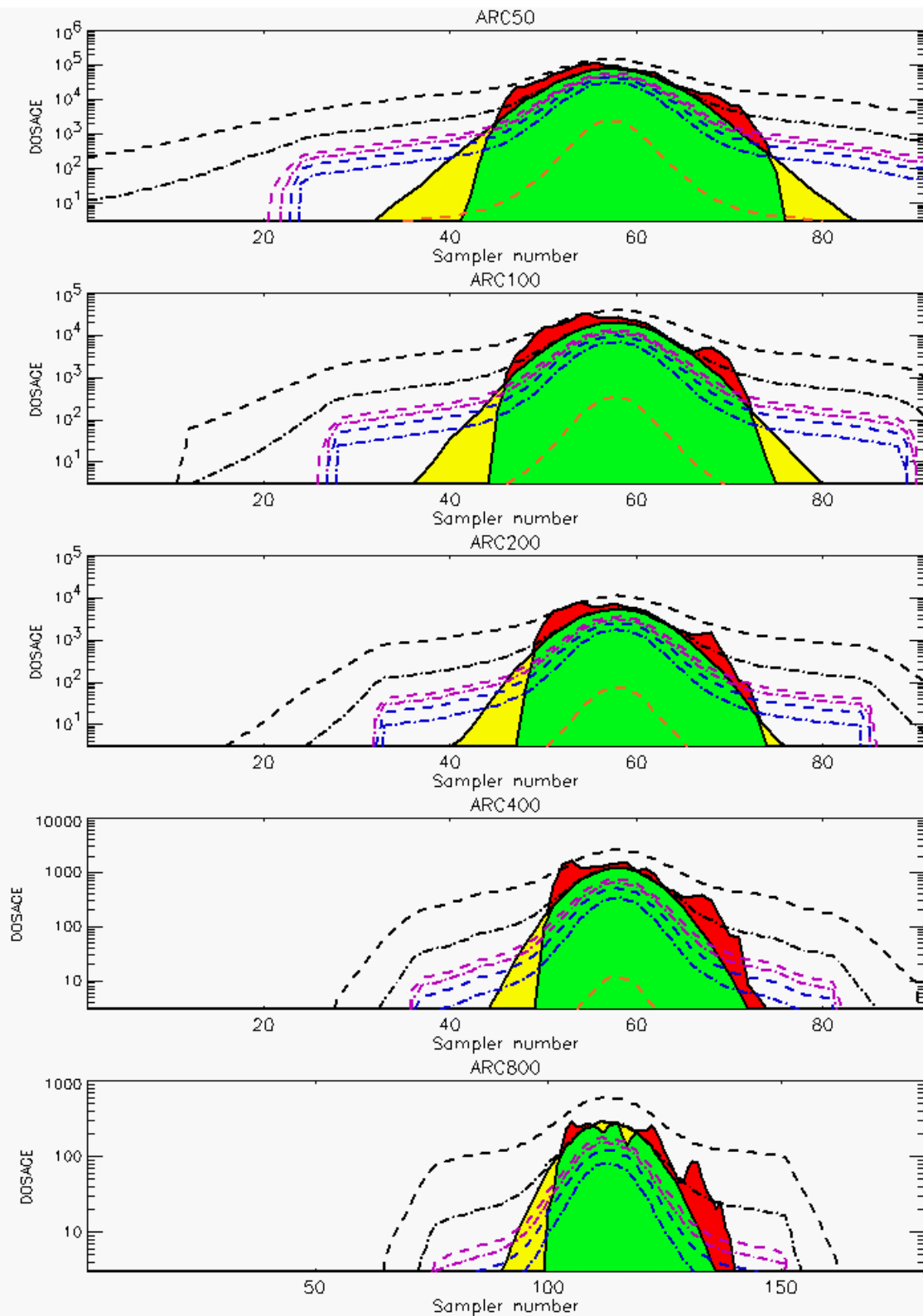


Figure C-4b. HPAC Probabilistic Prediction Outputs for Trial 9 on Logarithmic Scale: Stability Class is 2

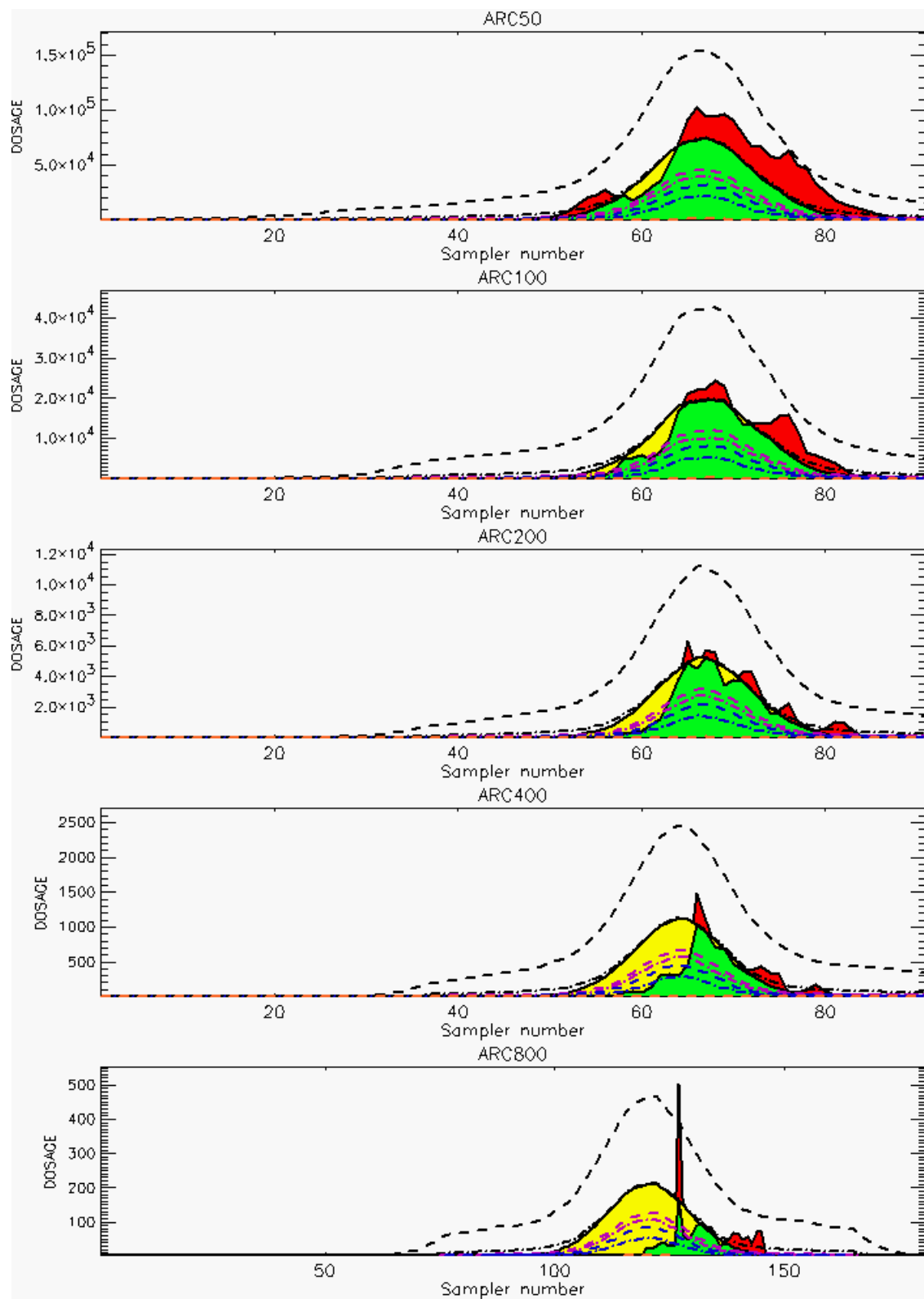


Figure C-5a. HPAC Probabilistic Prediction Outputs for Trial 10 on Linear Scale: Stability Class is 1

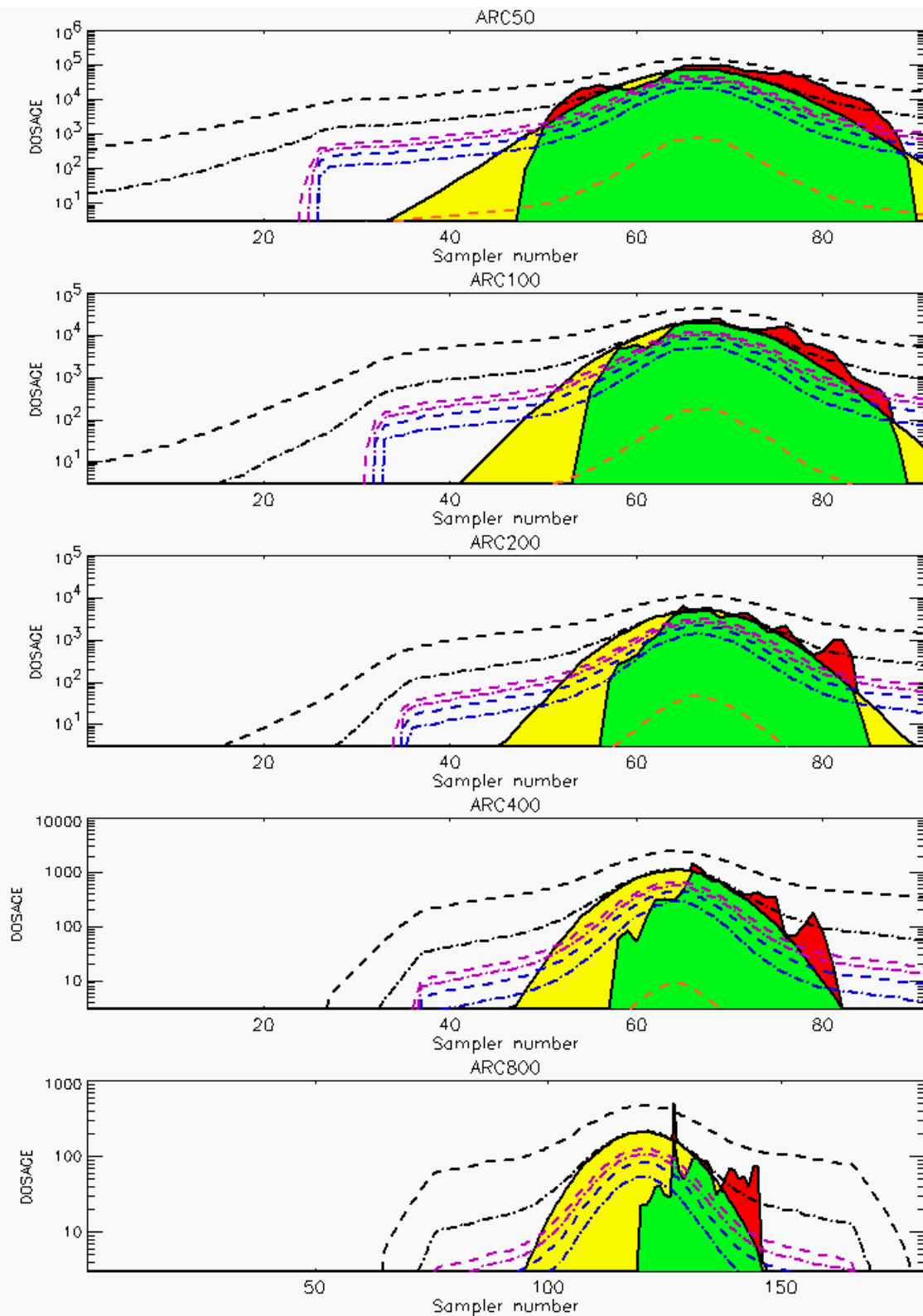


Figure C-5b. HPAC Probabilistic Prediction Outputs for Trial 10 on Logarithmic Scale: Stability Class is 1

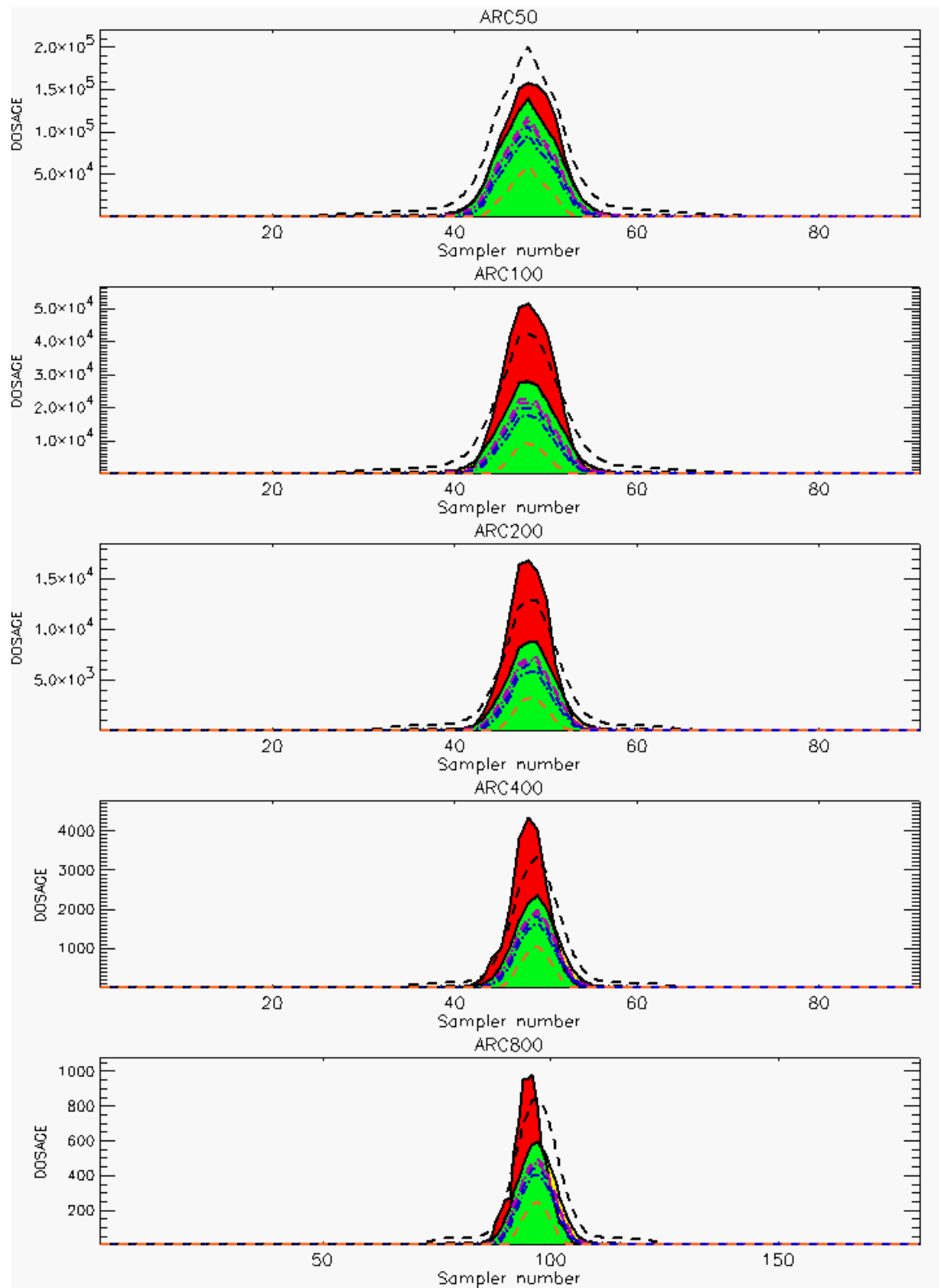


Figure C-6a. HPAC Probabilistic Prediction Outputs for Trial 11 on Linear Scale: Stability Class is 3 (Value for Sampler 90 of 800-Meter Arc is Considered “Spurious” and is Fixed by Moving Decimal Point)

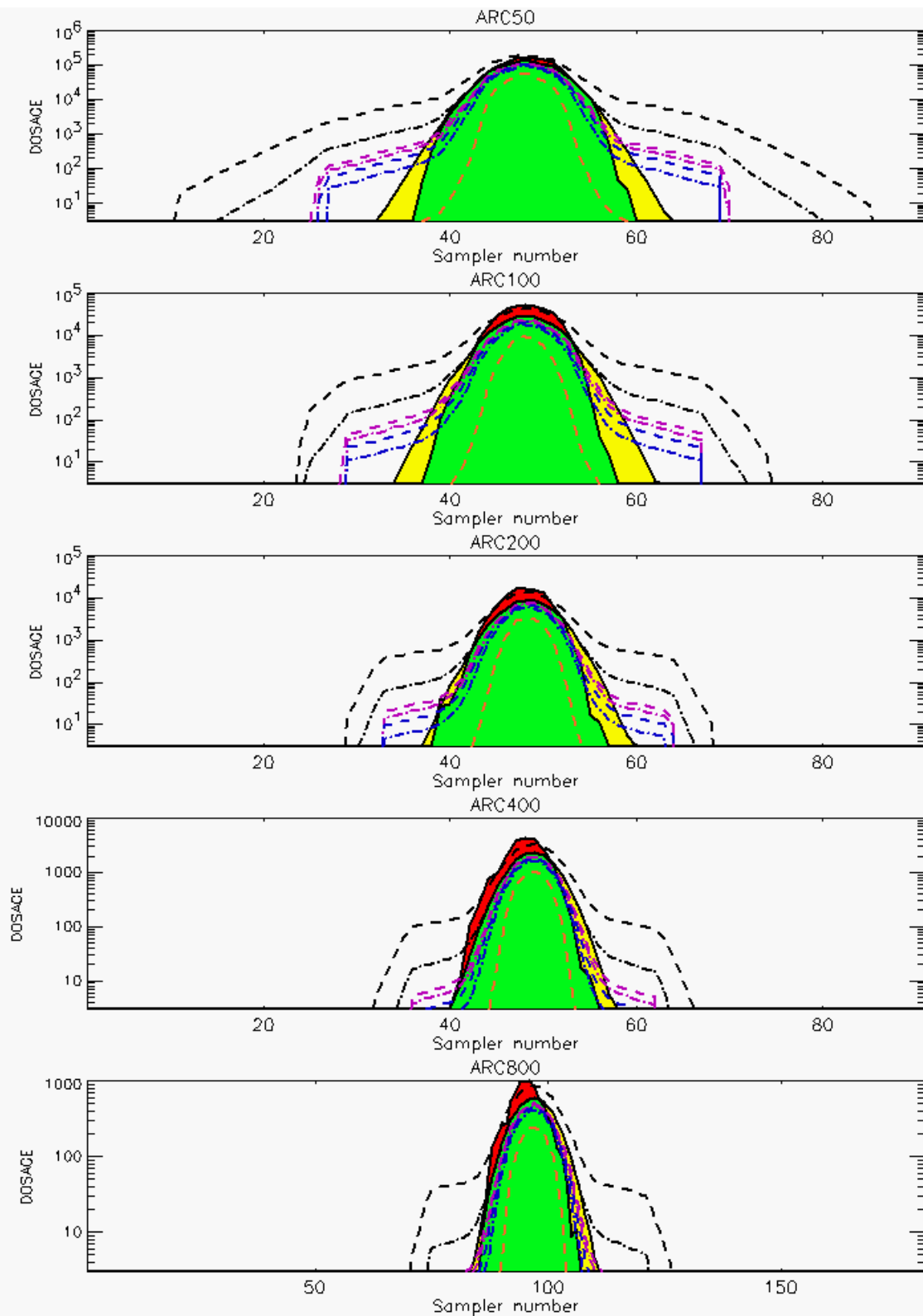


Figure C-6b. HPAC Probabilistic Prediction Outputs for Trial 11 on Logarithmic Scale: Stability Class is 3 (Value for Sampler 90 of 800-Meter Arc is Considered “Spurious” and is Fixed by Moving Decimal Point)

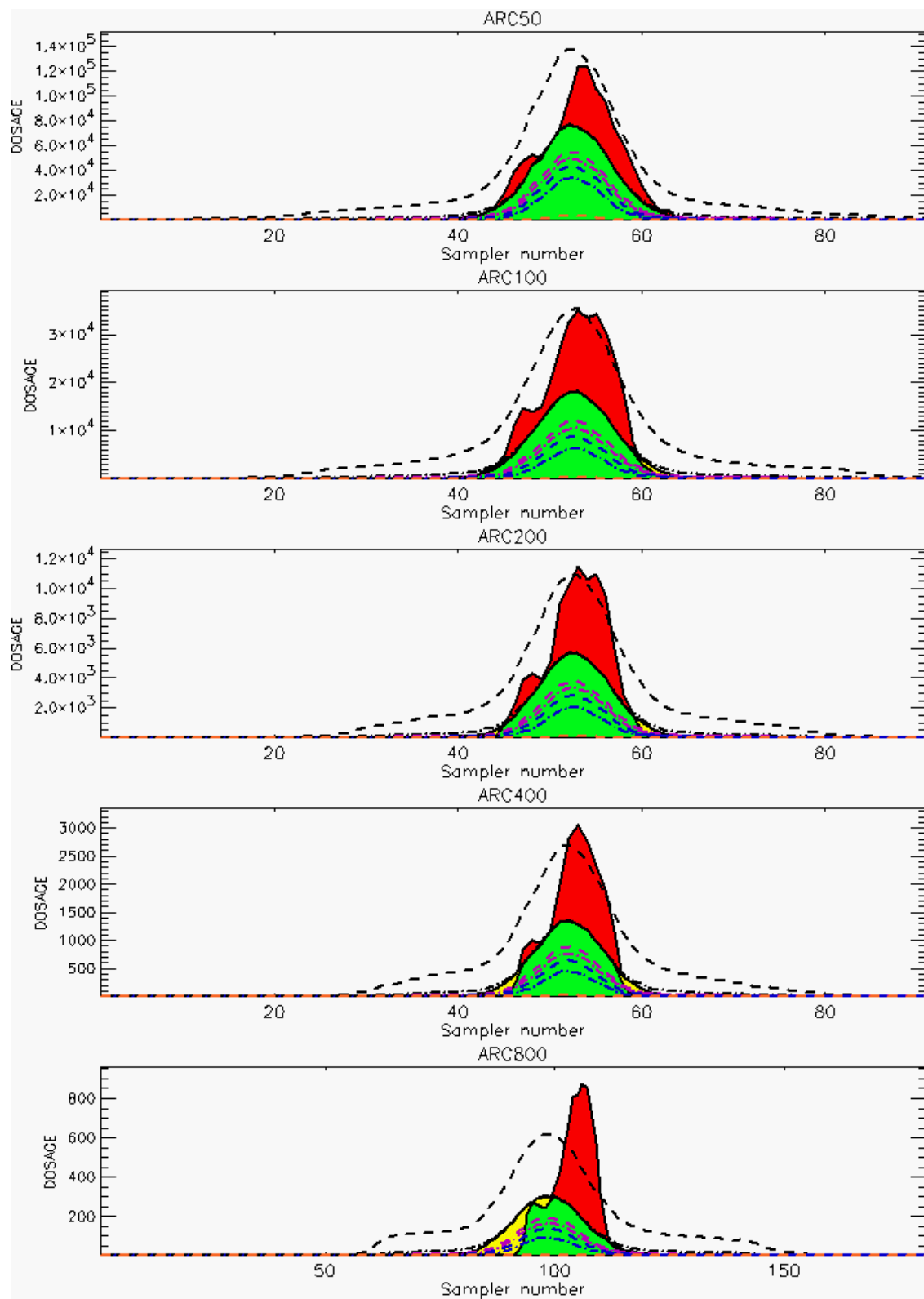
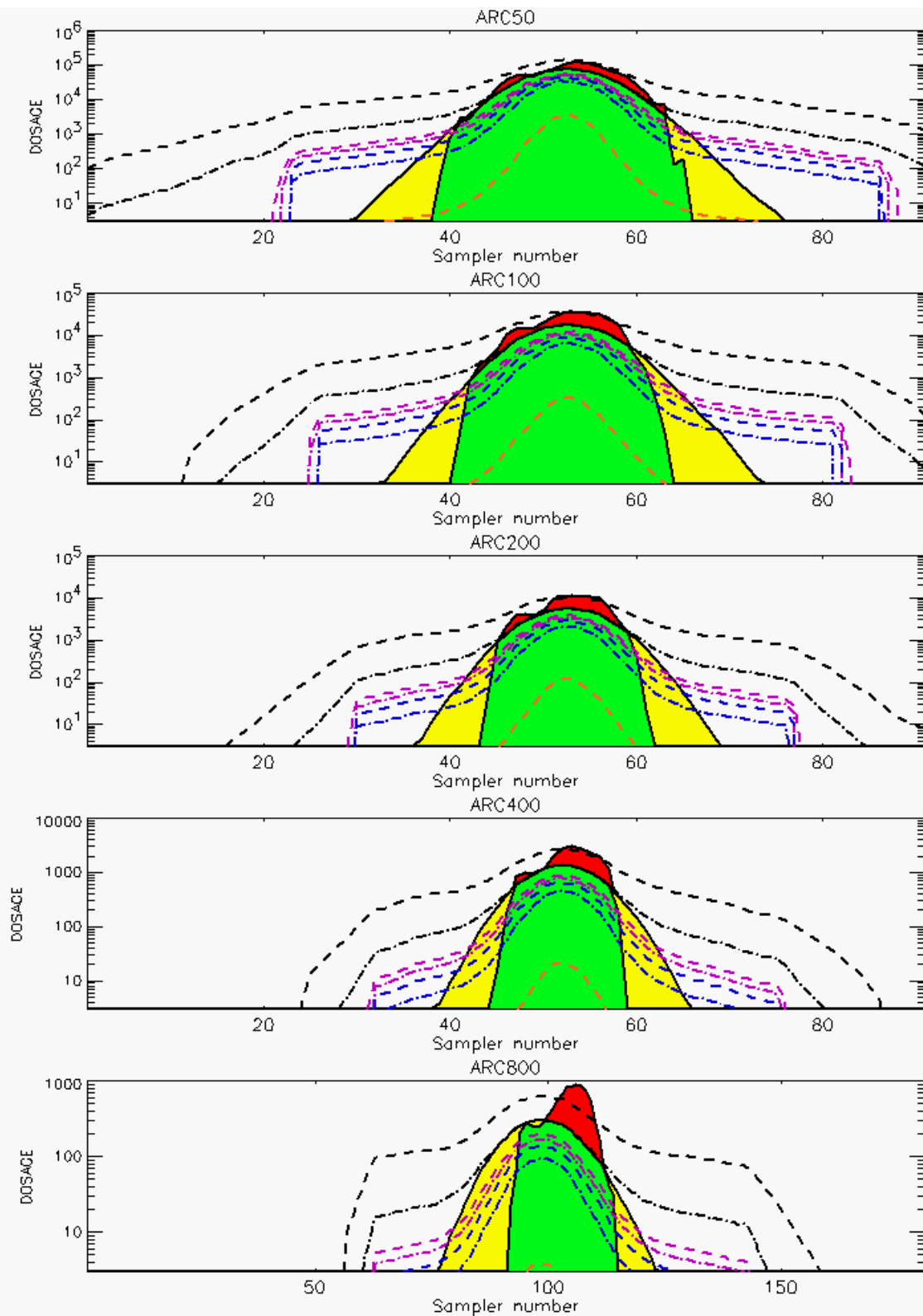


Figure C-7a. HPAC Probabilistic Prediction Outputs for Trial 12 on Linear Scale: Stability Class is 3



**Figure C-7b. HPAC Probabilistic Prediction Outputs for Trial 12 on Logarithmic Scale:
Stability Class is 3**

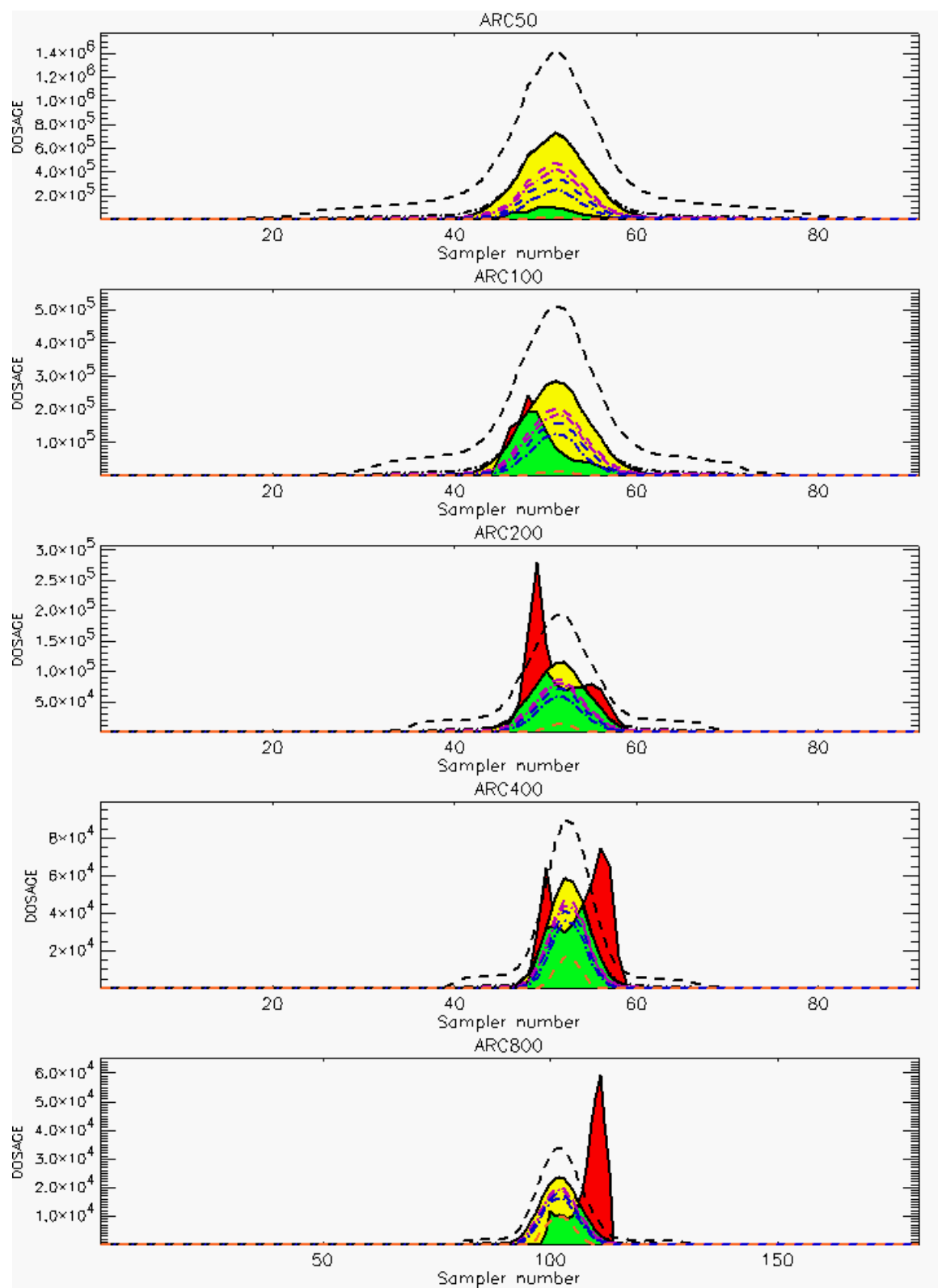
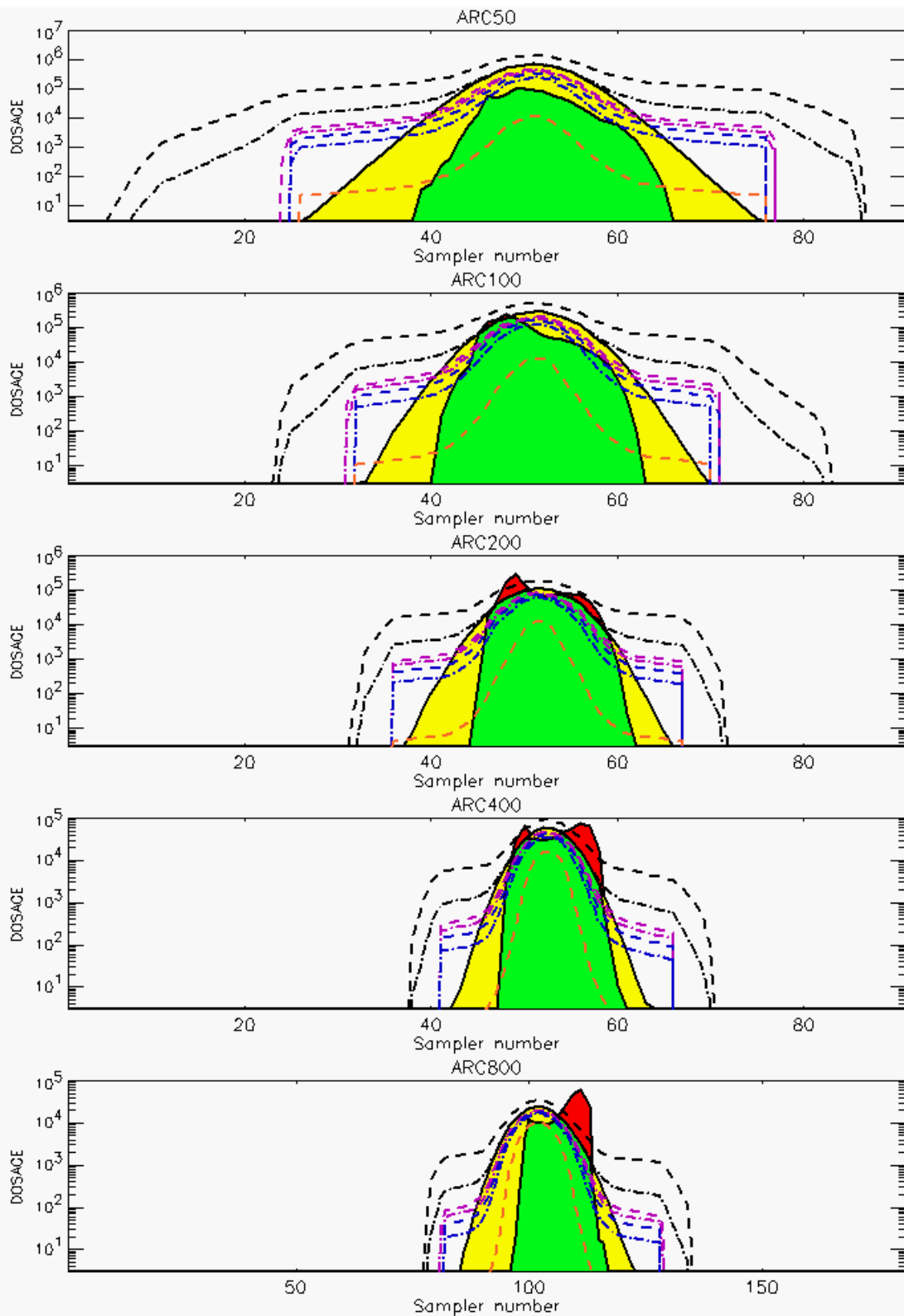


Figure C-8a. HPAC Probabilistic Prediction Outputs for Trial 13 on Linear Scale: Stability Class is 7



**Figure C-8b. HPAC Probabilistic Prediction Outputs for Trial 13 on Logarithmic Scale:
Stability Class is 7**

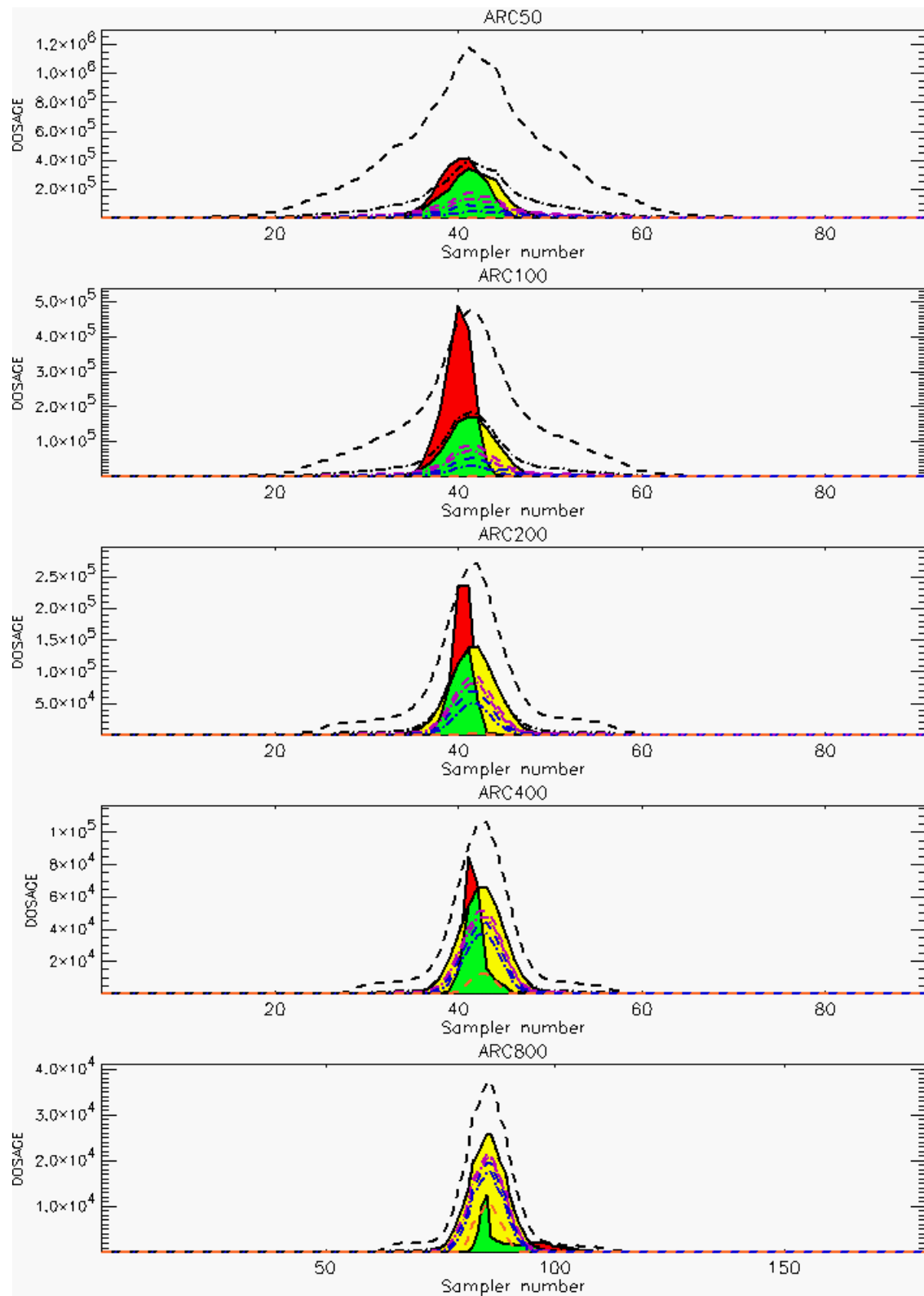
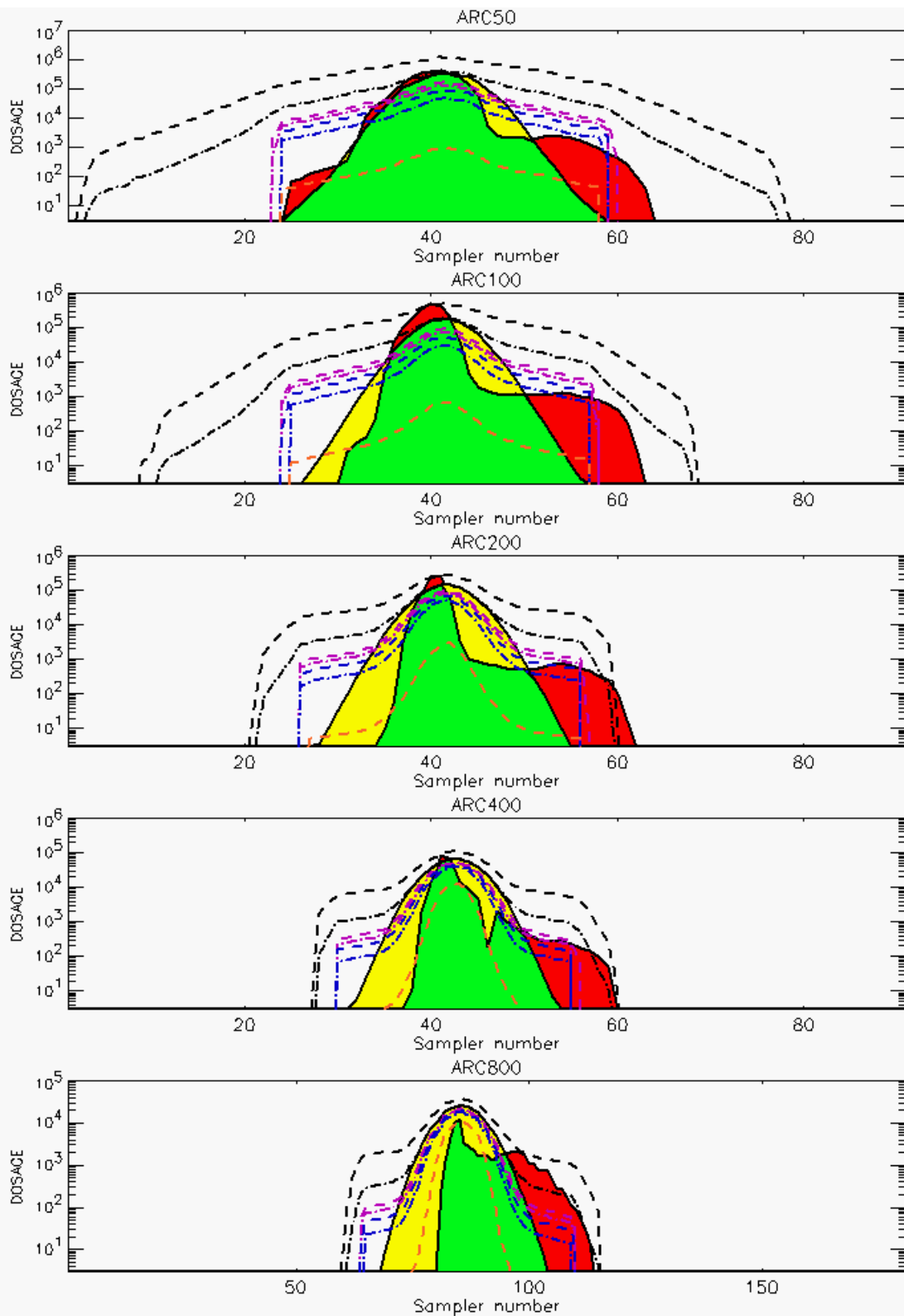


Figure C-9a. HPAC Probabilistic Prediction Outputs for Trial 14 on Linear Scale: Stability Class is 7



**Figure C-9b. HPAC Probabilistic Prediction Outputs for Trial 14 on Logarithmic Scale:
Stability Class is 7**

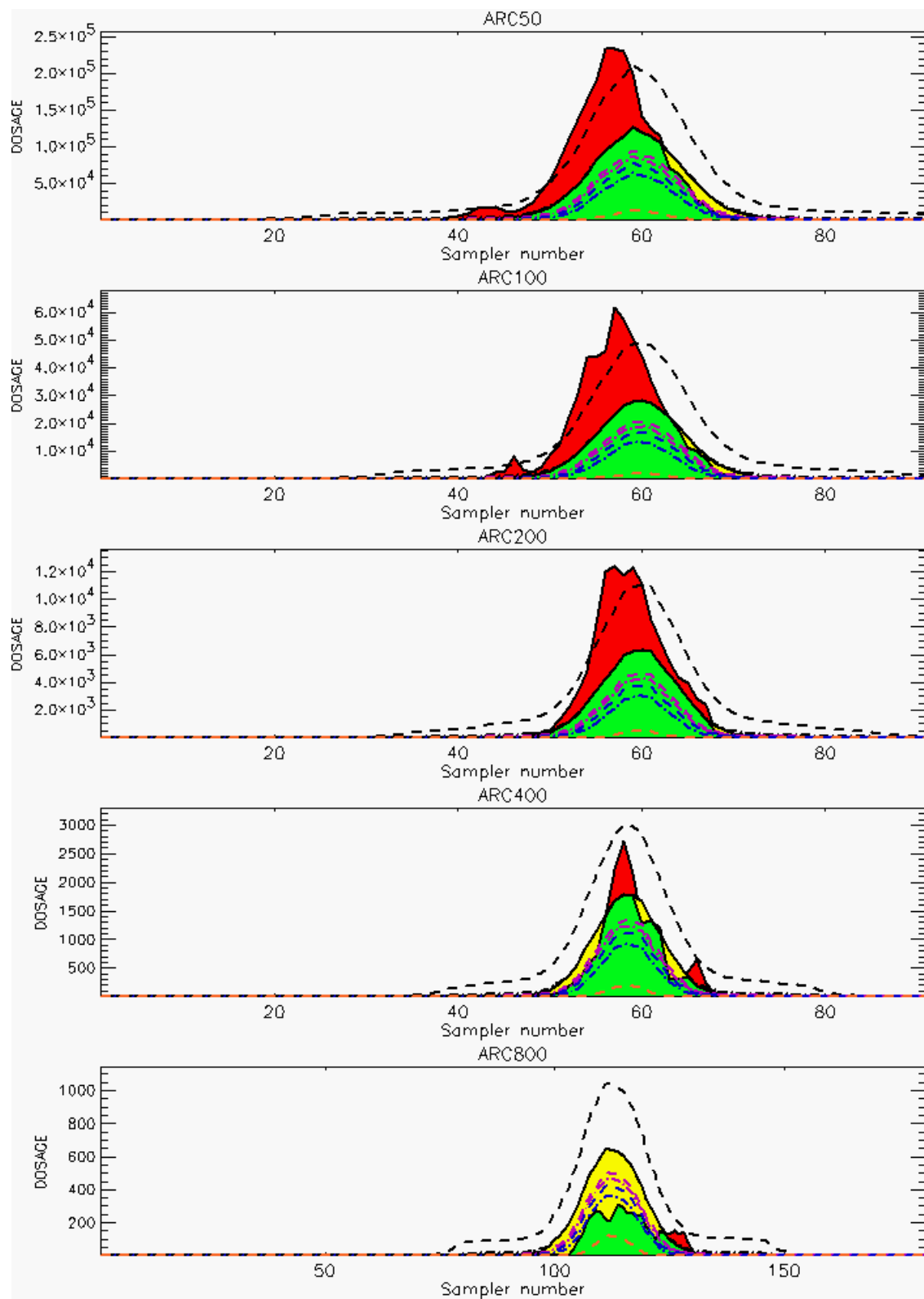


Figure C-10a. HPAC Probabilistic Prediction Outputs for Trial 15 on Linear Scale: Stability Class is 1

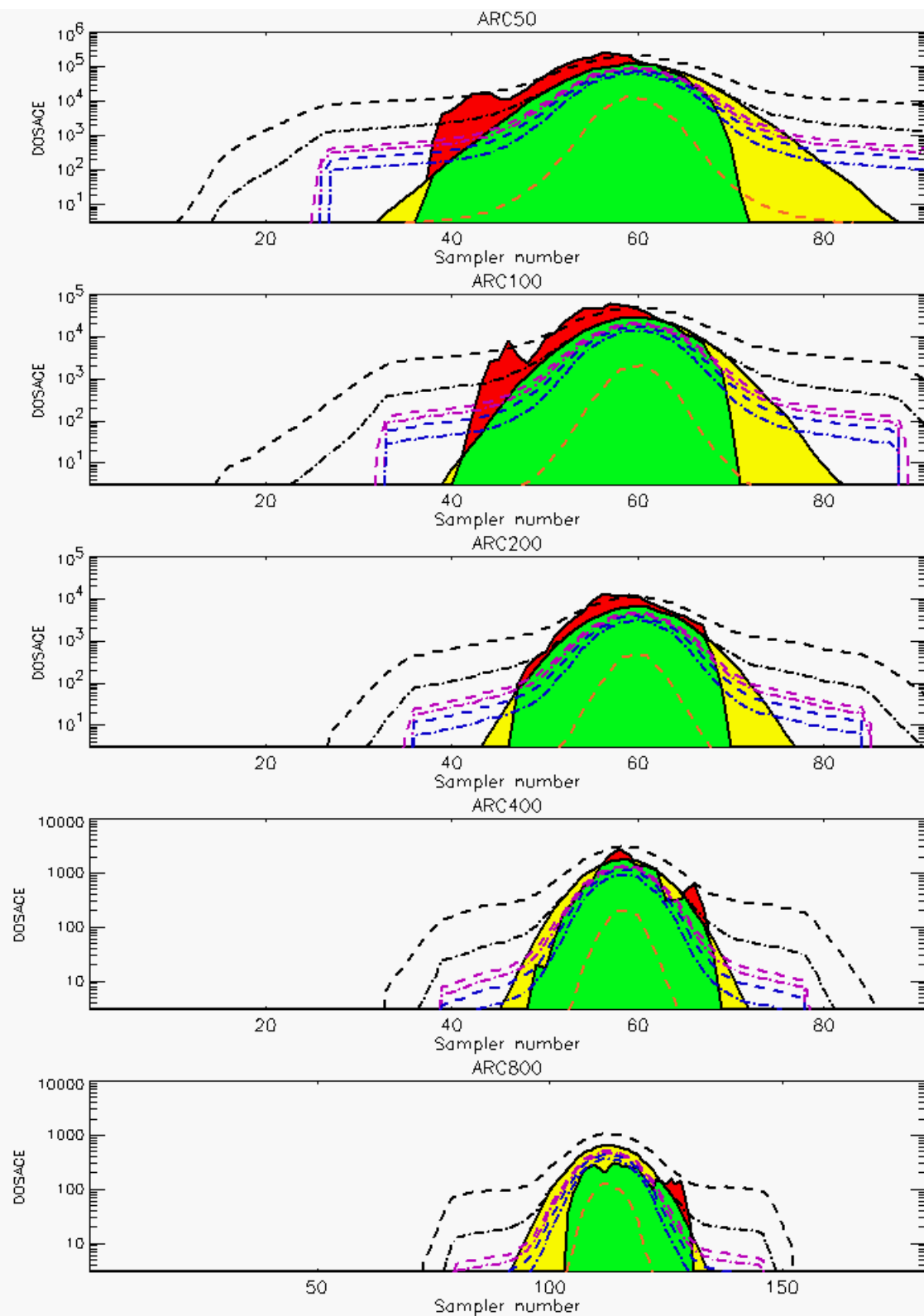


Figure C-10b. HPAC Probabilistic Prediction Outputs for Trial 15 on Logarithmic Scale: Stability Class is 1

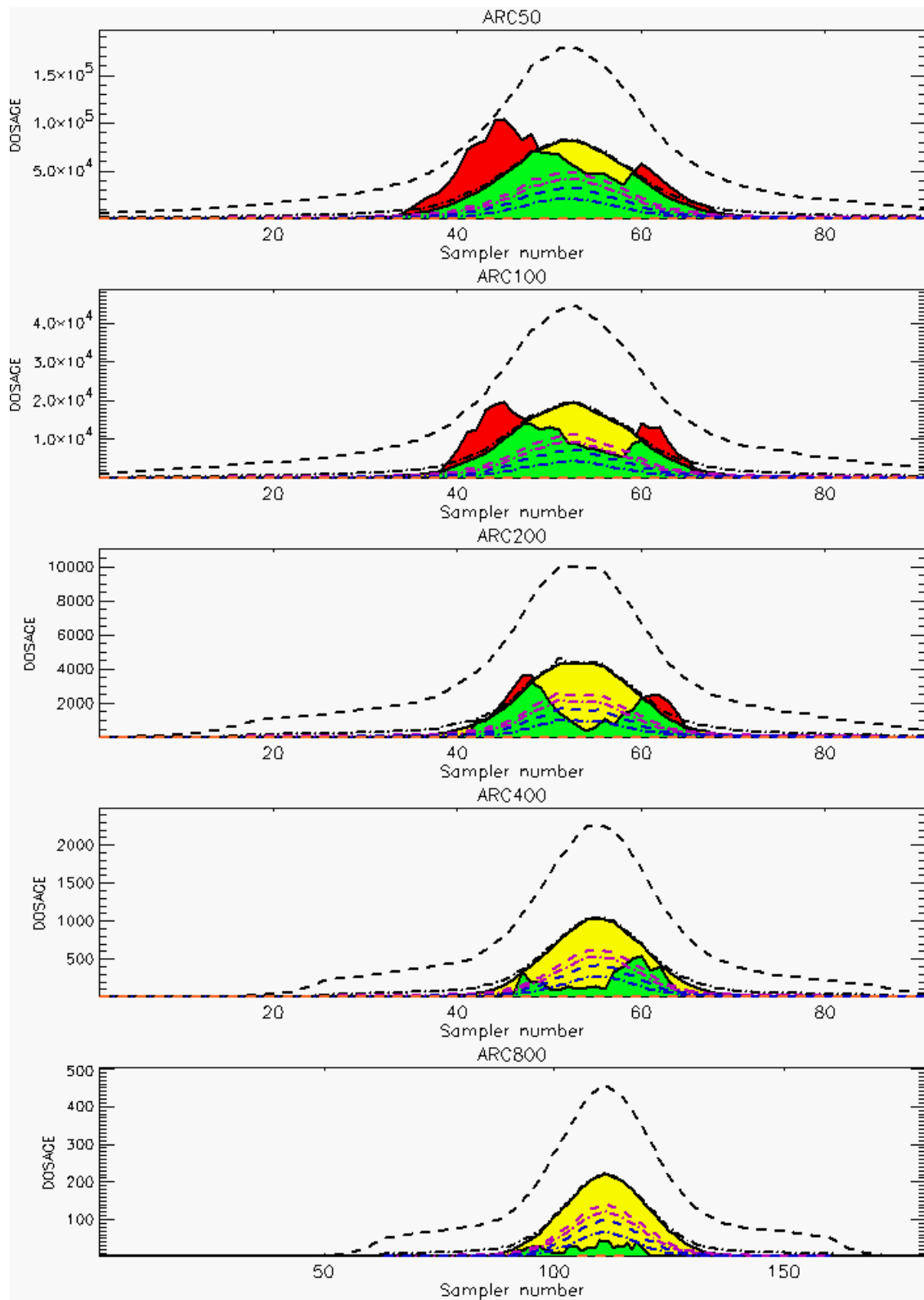


Figure C-11a. HPAC Probabilistic Prediction Outputs for Trial 16 on Linear Scale: Stability Class is 1 (Value for Sampler 63 of 200-Meter Arc is Considered “Spurious” and is Fixed by Moving Decimal Point)

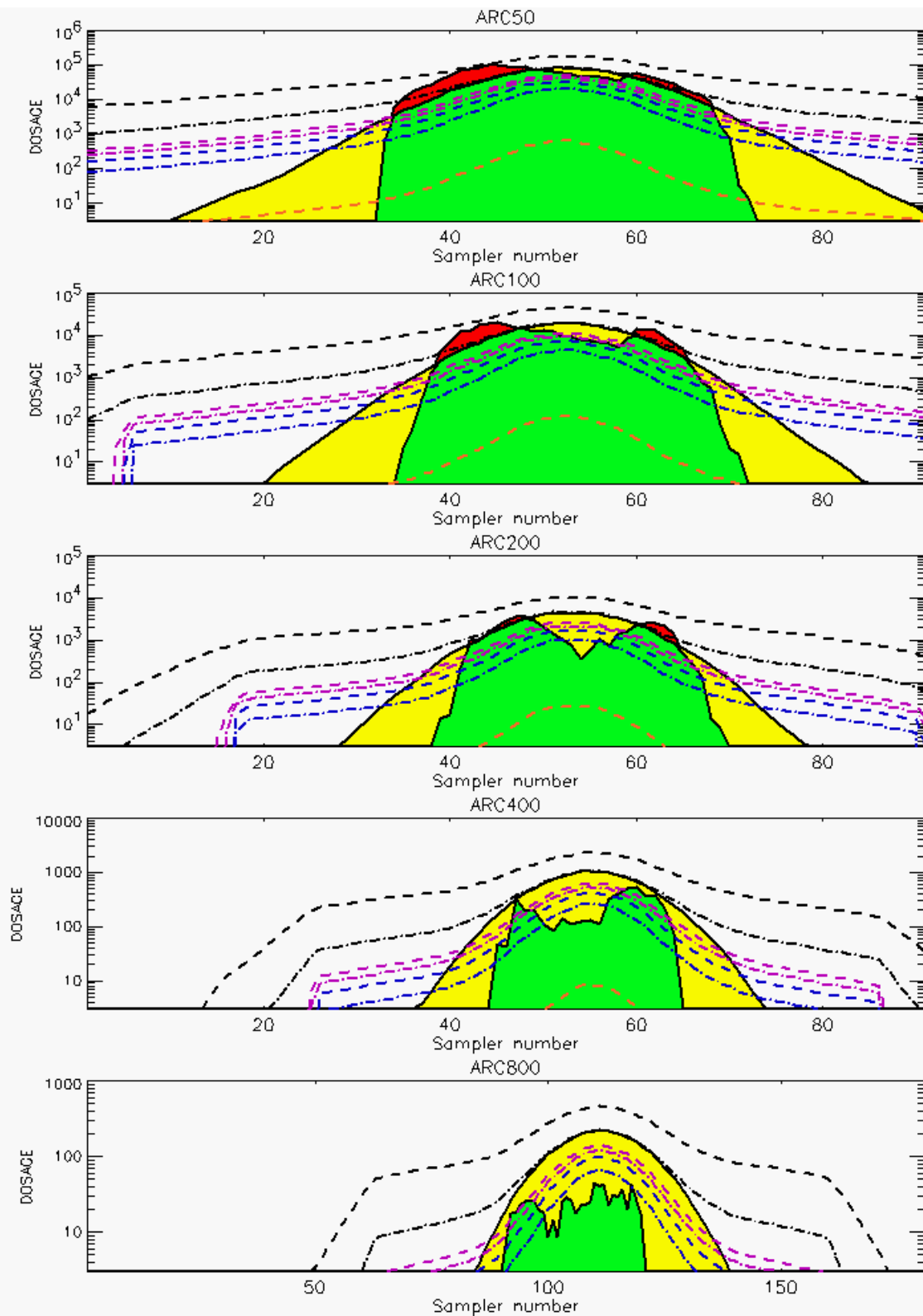


Figure C-11b. HPAC Probabilistic Prediction Outputs for Trial 16 on Logarithmic Scale: Stability Class is 1 (Value for Sampler 63 of 200-Meter Arc is Considered “Spurious” and is Fixed by Moving Decimal Point)

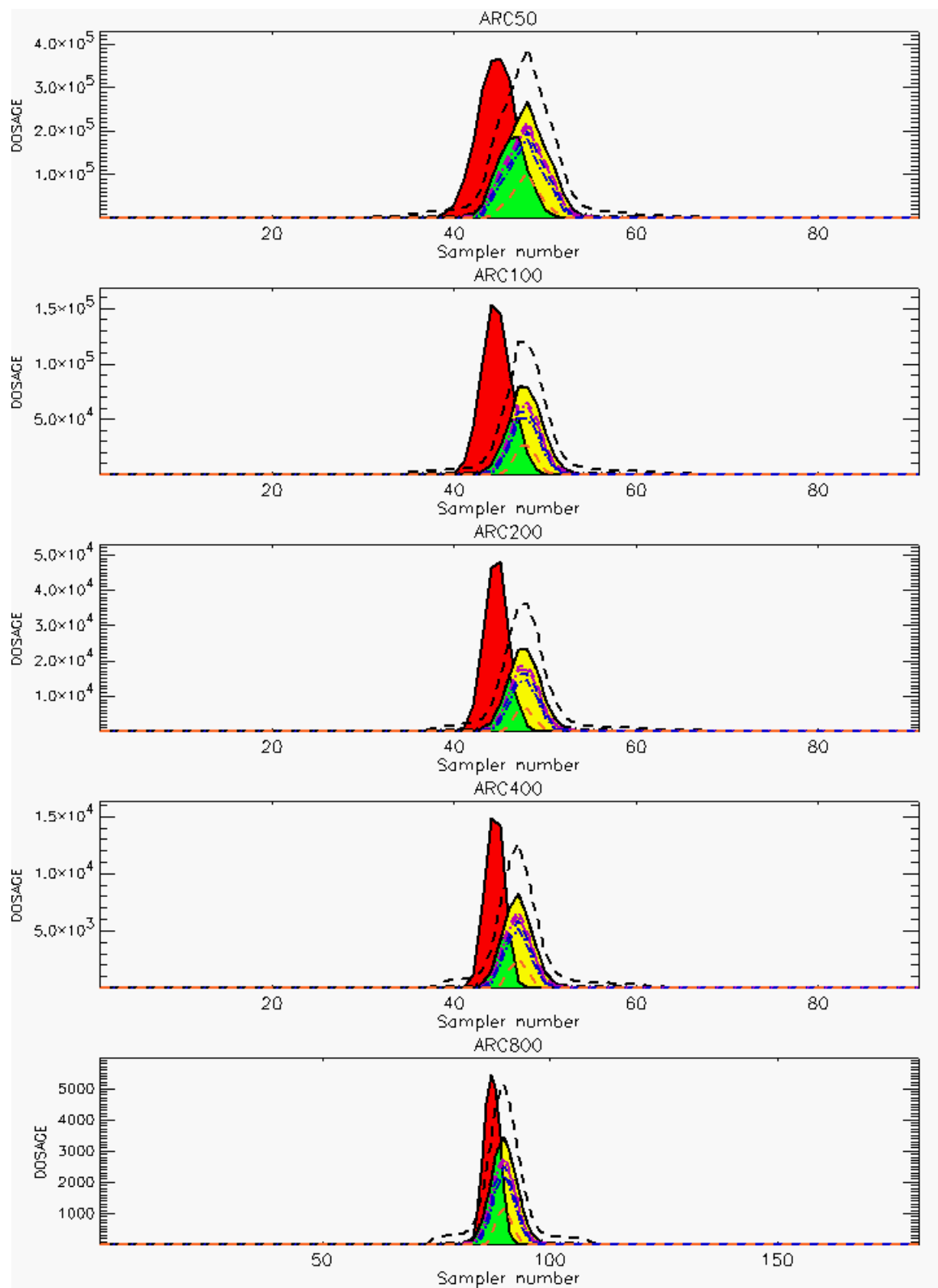


Figure C-12a. HPAC Probabilistic Prediction Outputs for Trial 17 on Linear Scale: Stability Class is 5

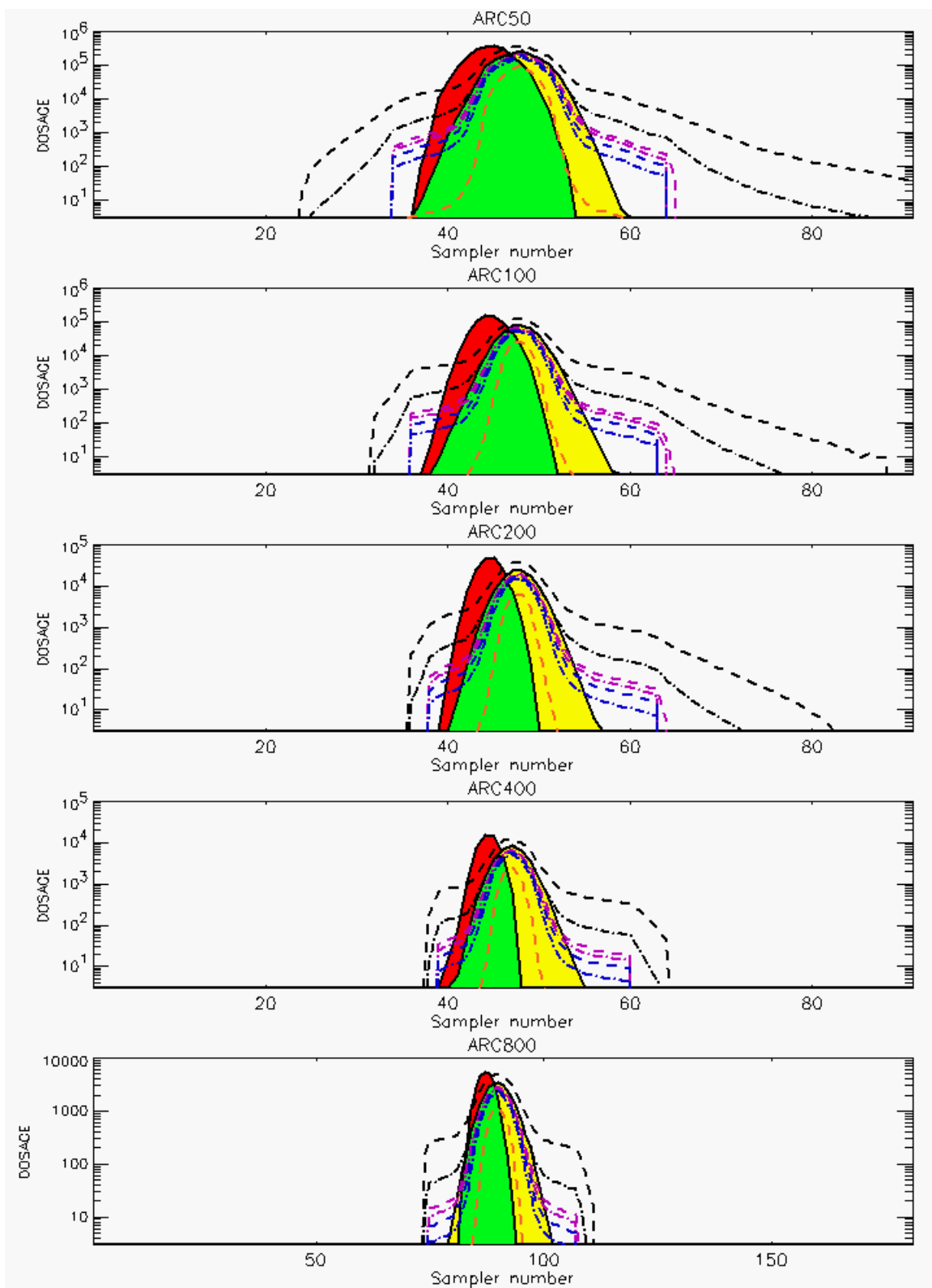


Figure C-12b. HPAC Probabilistic Prediction Outputs for Trial 17 on Logarithmic Scale: Stability Class is 5

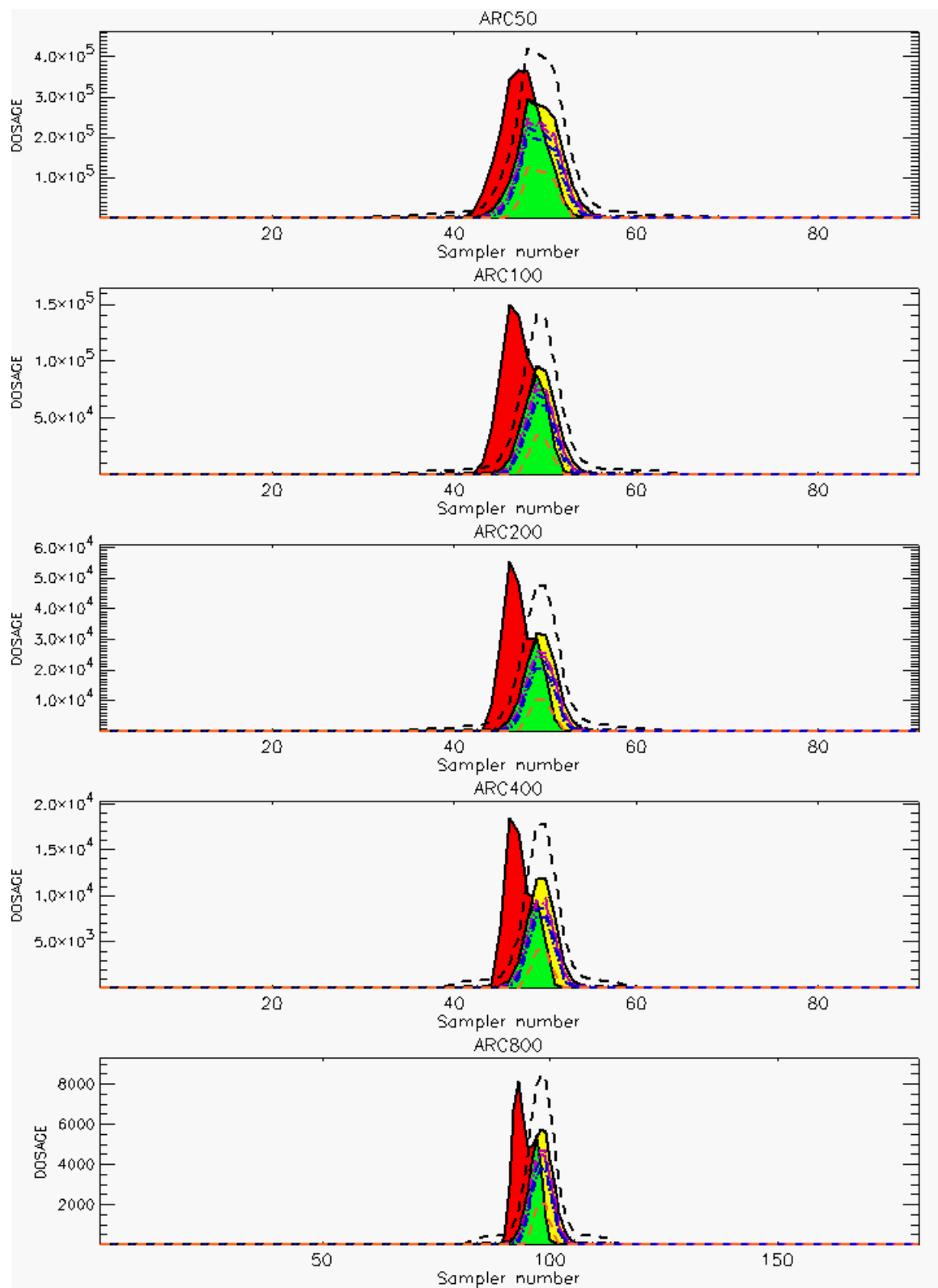


Figure C-13a. HPAC Probabilistic Prediction Outputs for Trial 18 on Linear Scale: Stability Class is 5

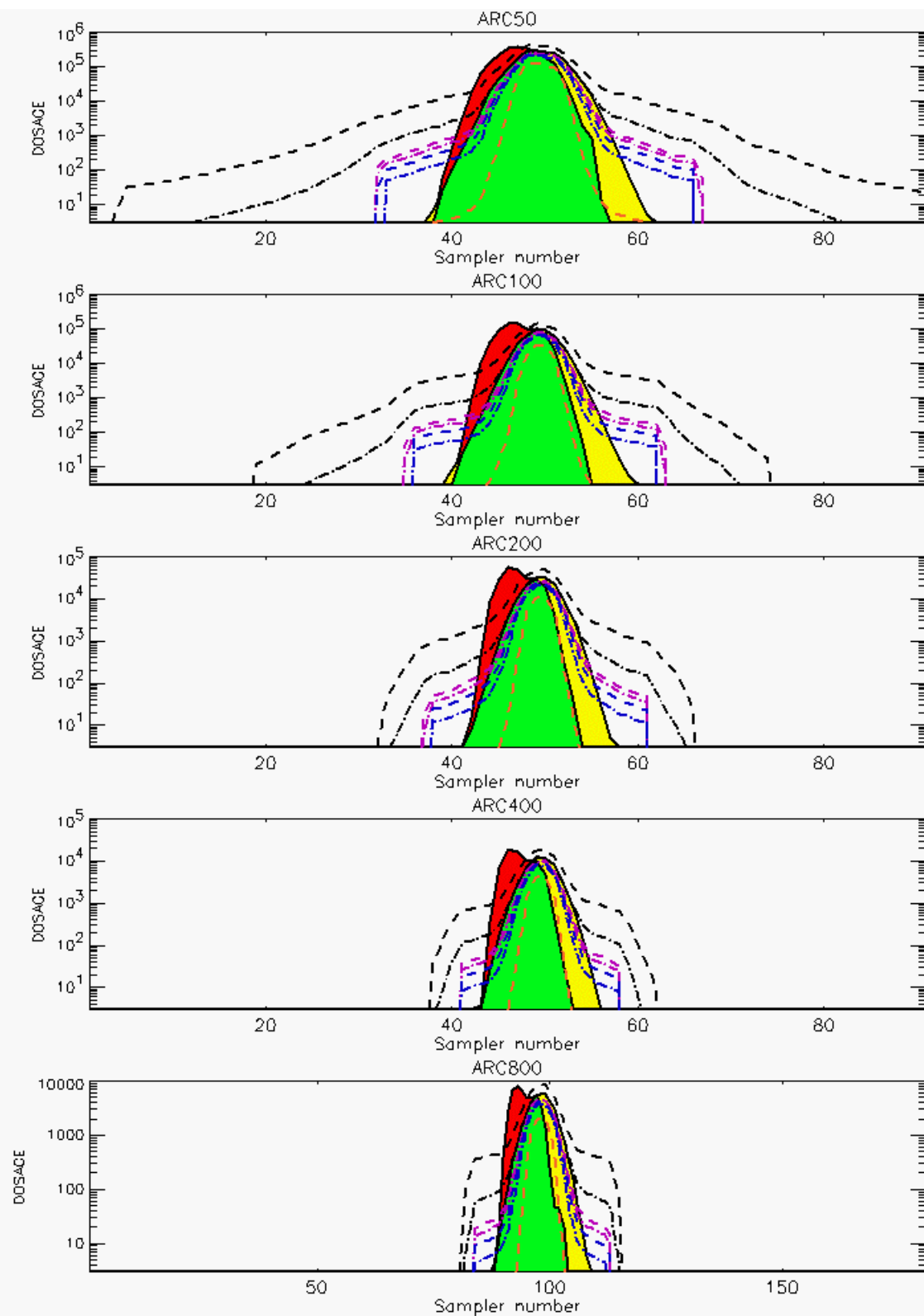


Figure C-13b. HPAC Probabilistic Prediction Outputs for Trial 18 on Logarithmic Scale: Stability Class is 5

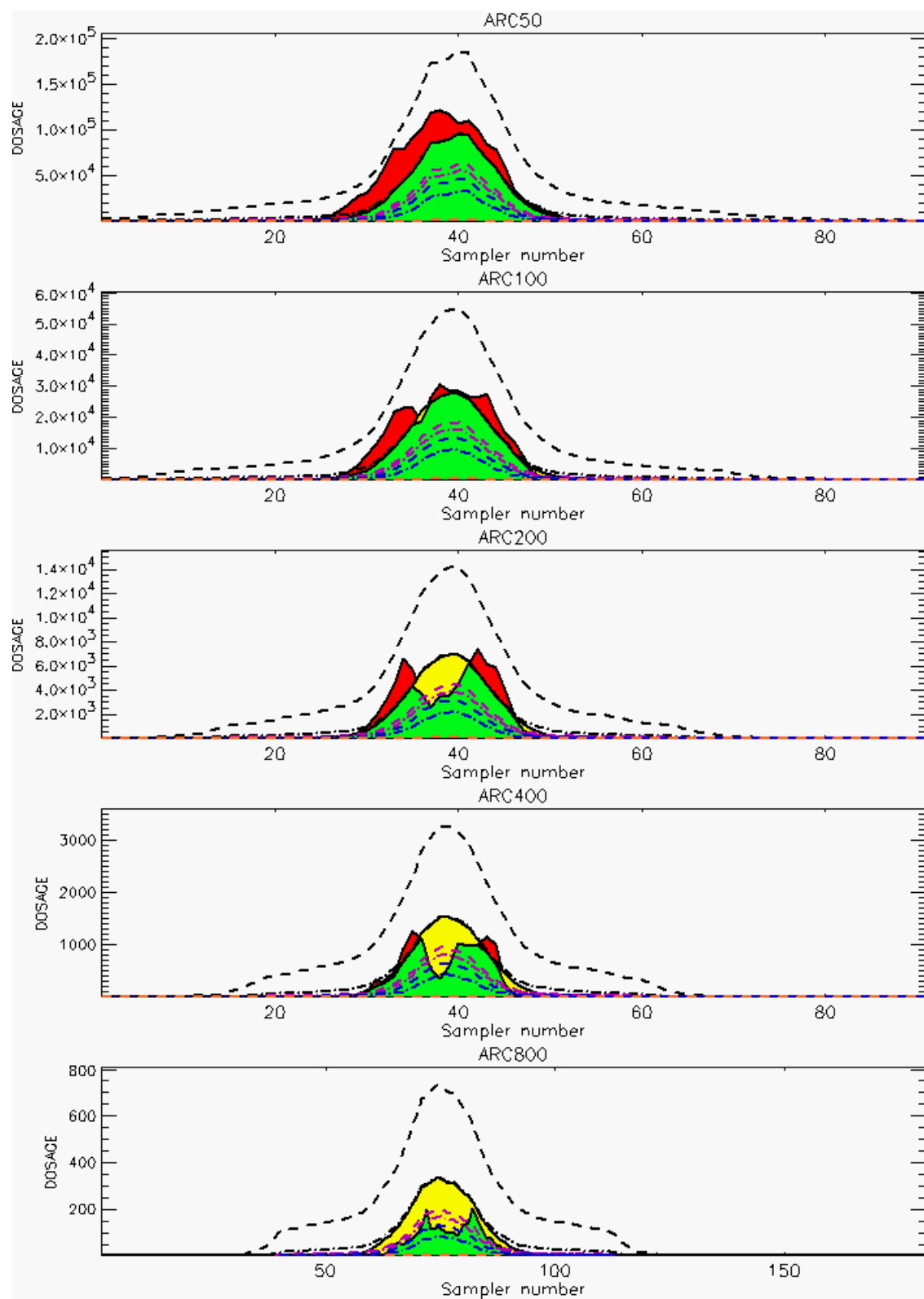


Figure C-14a. HPAC Probabilistic Prediction Outputs for Trial 19 on Linear Scale: Stability Class is 2

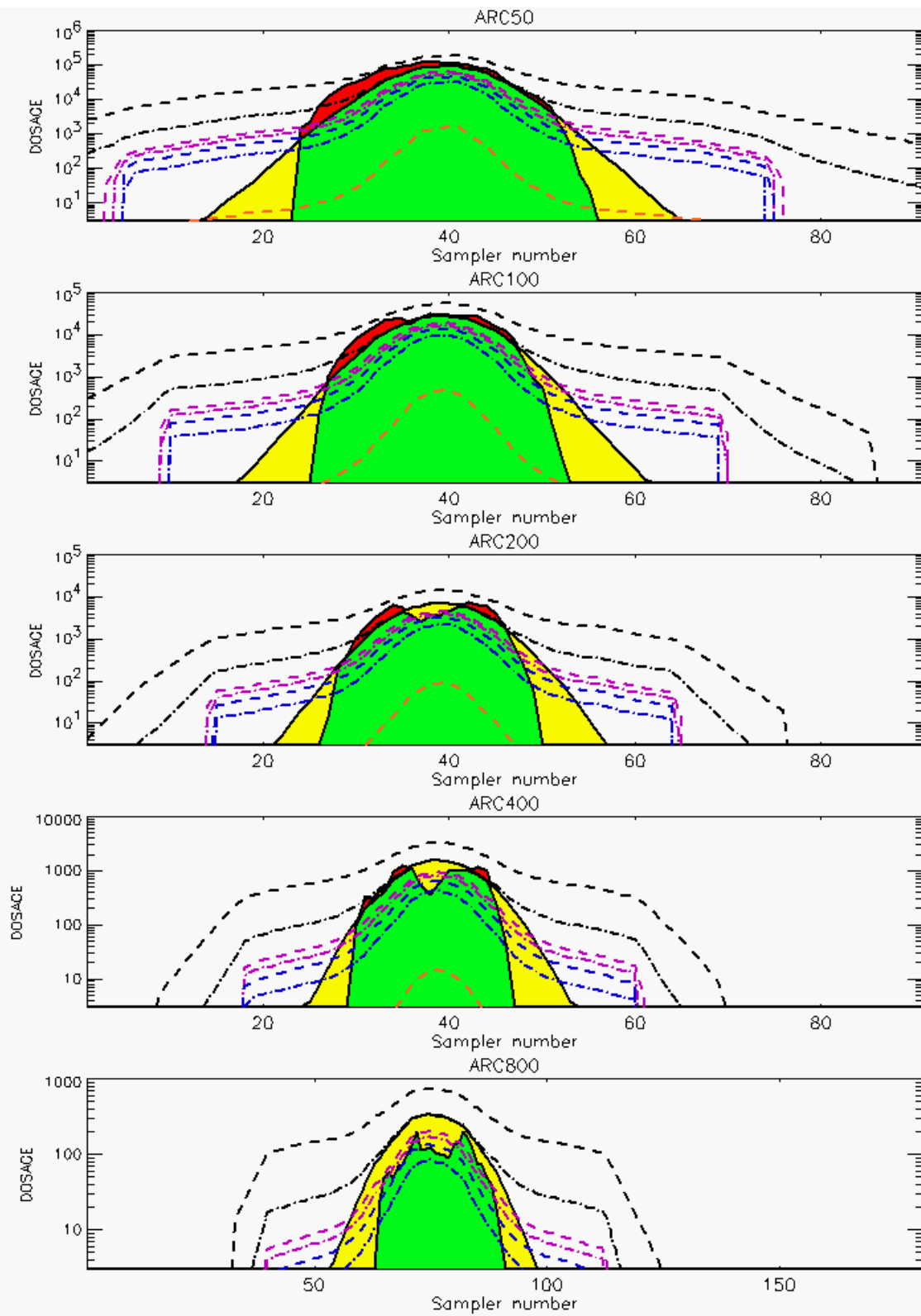


Figure C-14b. HPAC Probabilistic Prediction Outputs for Trial 19 on Logarithmic Scale: Stability Class is 2

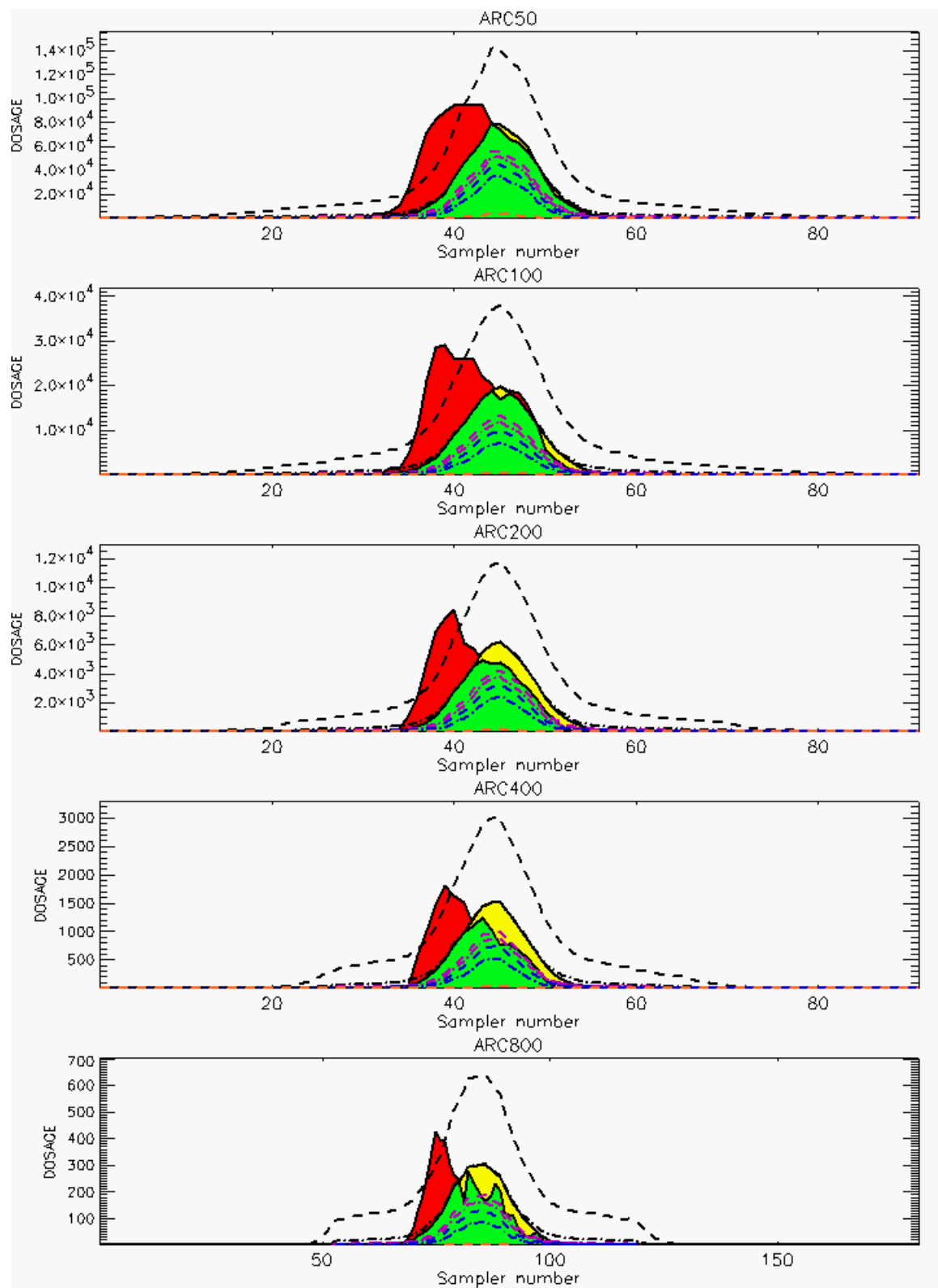


Figure C-15a. HPAC Probabilistic Prediction Outputs for Trial 20 on Linear Scale: Stability Class is 3

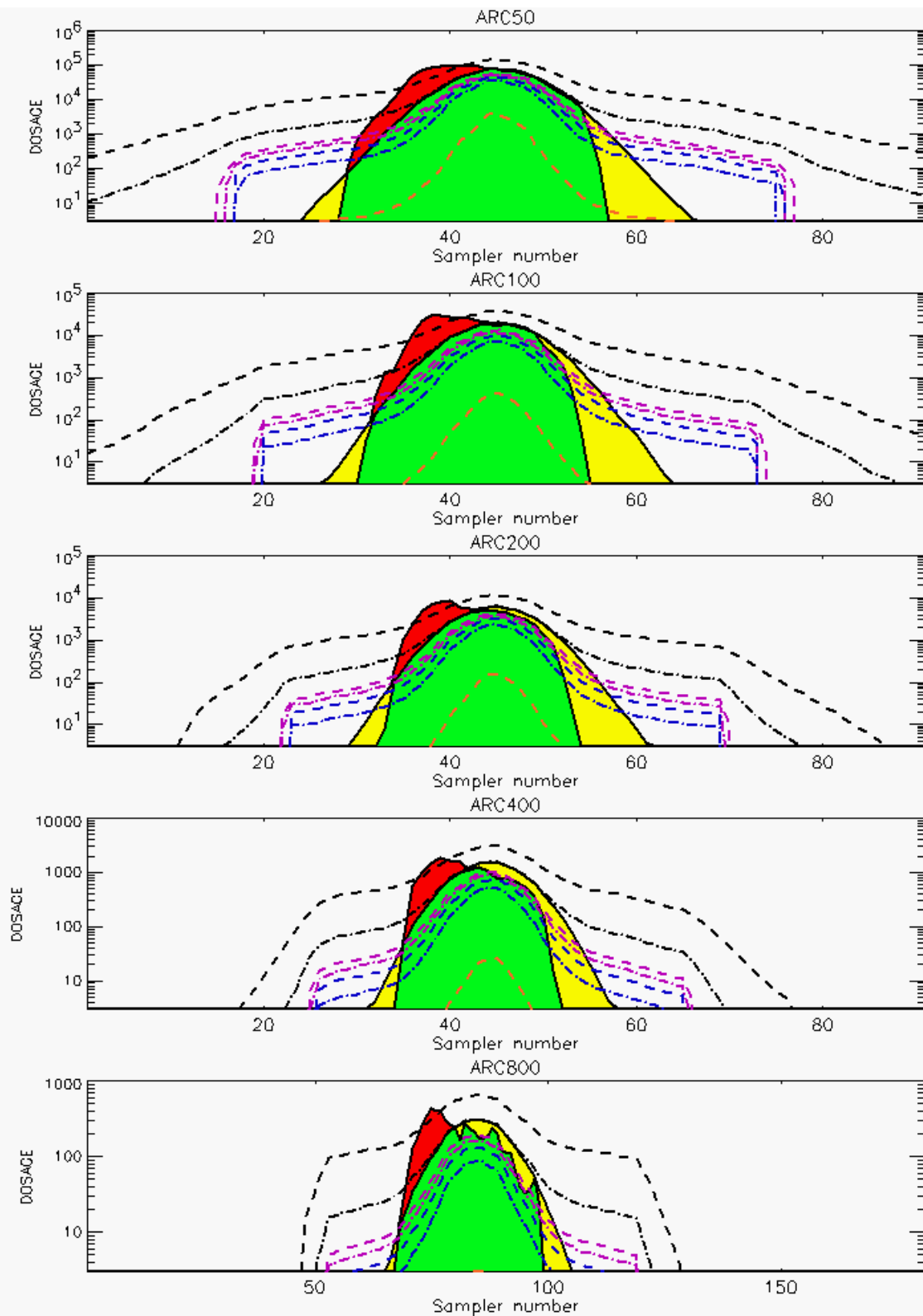


Figure C-15b. HPAC Probabilistic Prediction Outputs for Trial 20 on Logarithmic Scale: Stability Class is 3

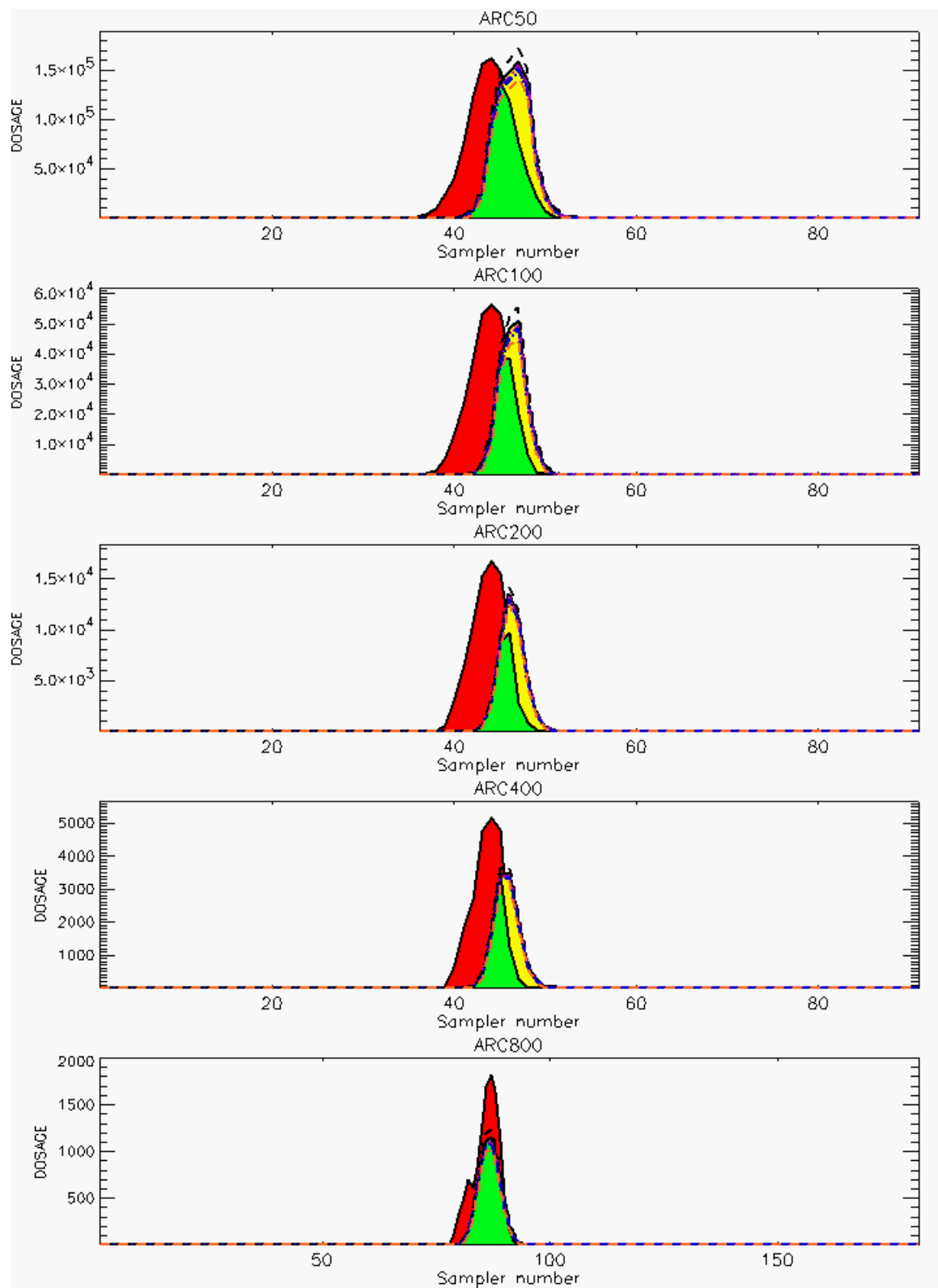


Figure C-16a. HPAC Probabilistic Prediction Outputs for Trial 21 on Linear Scale: Stability Class is 4

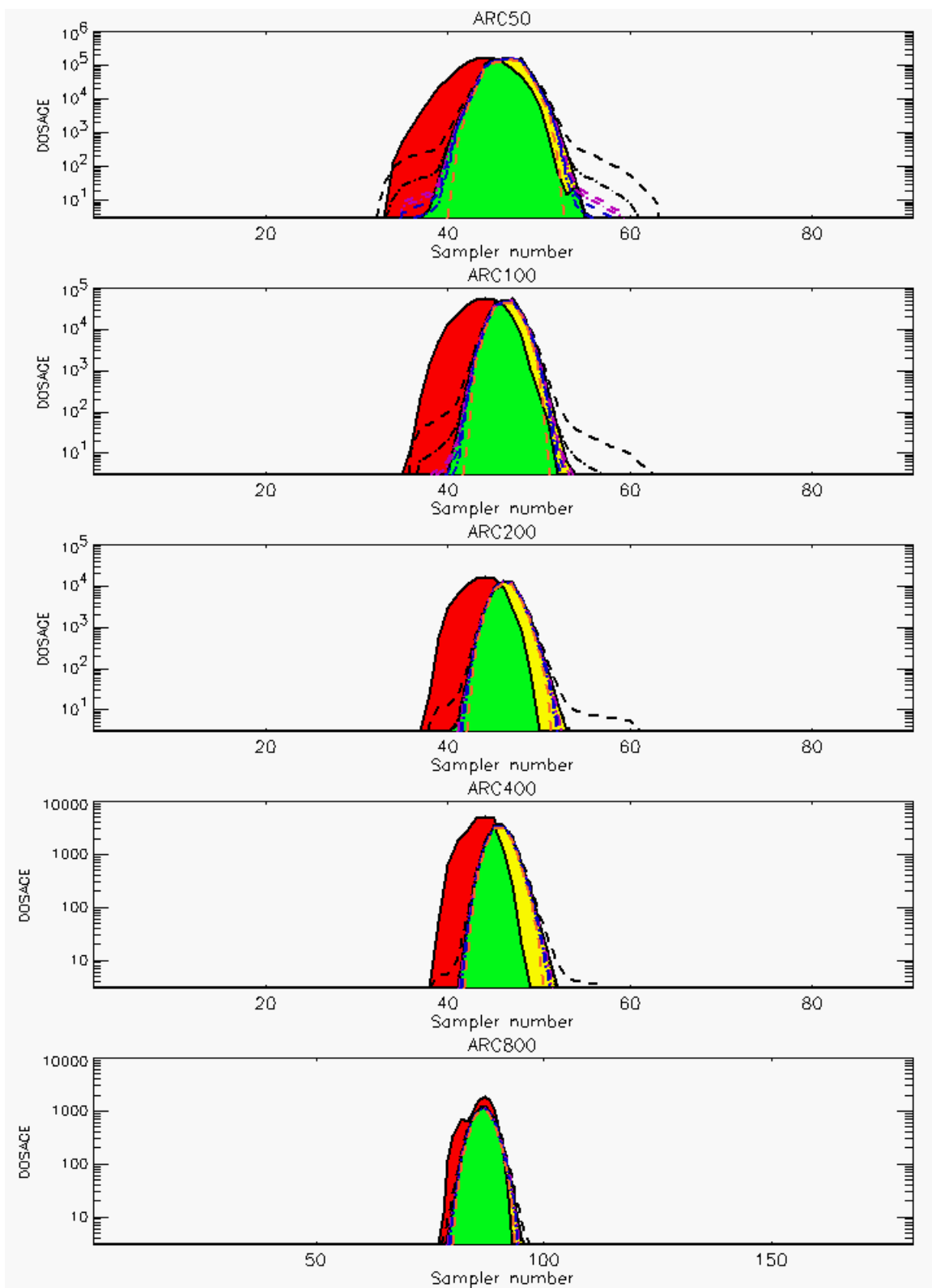


Figure C-16b. HPAC Probabilistic Prediction Outputs for Trial 21 on Logarithmic Scale: Stability Class is 4

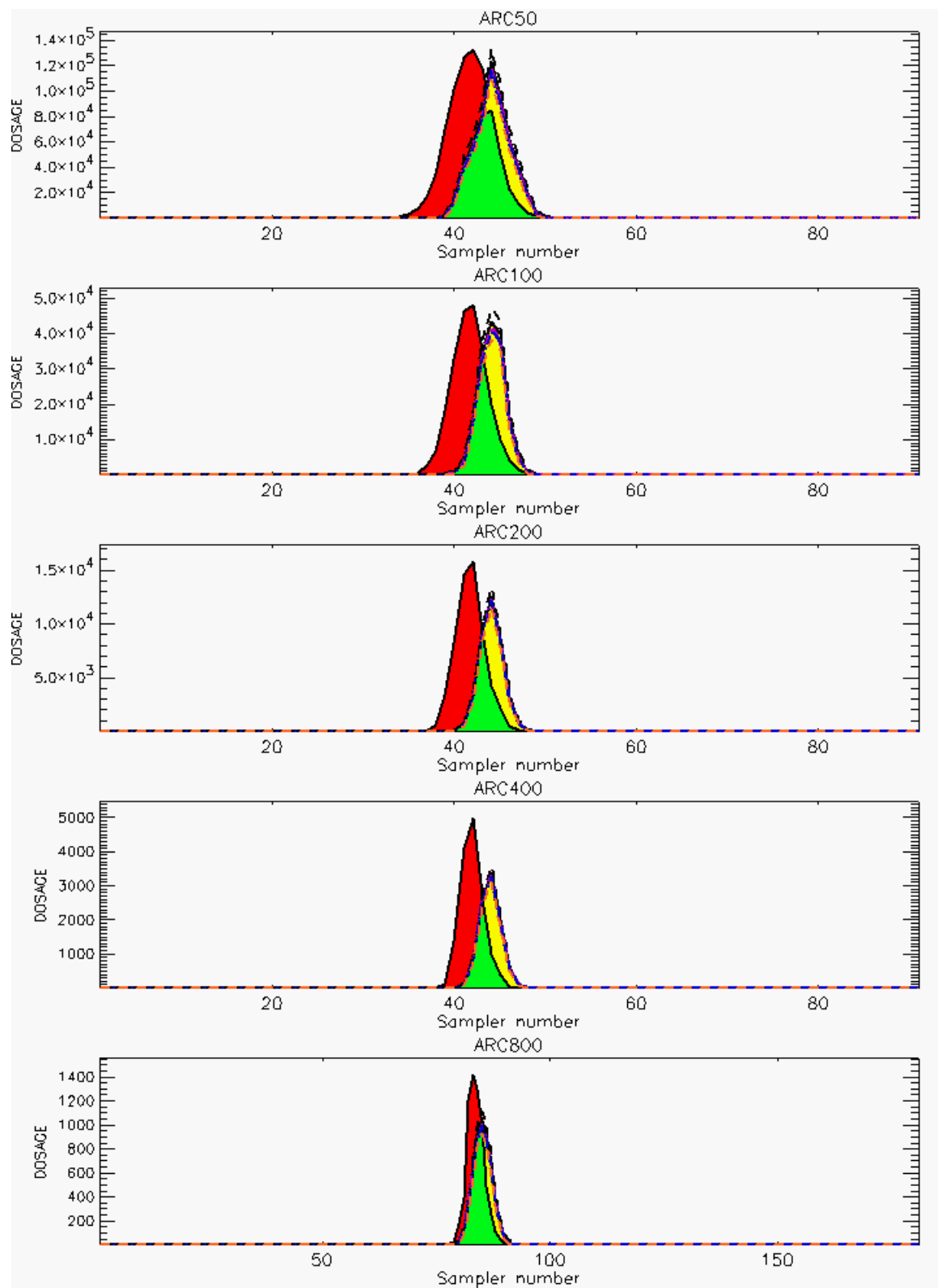
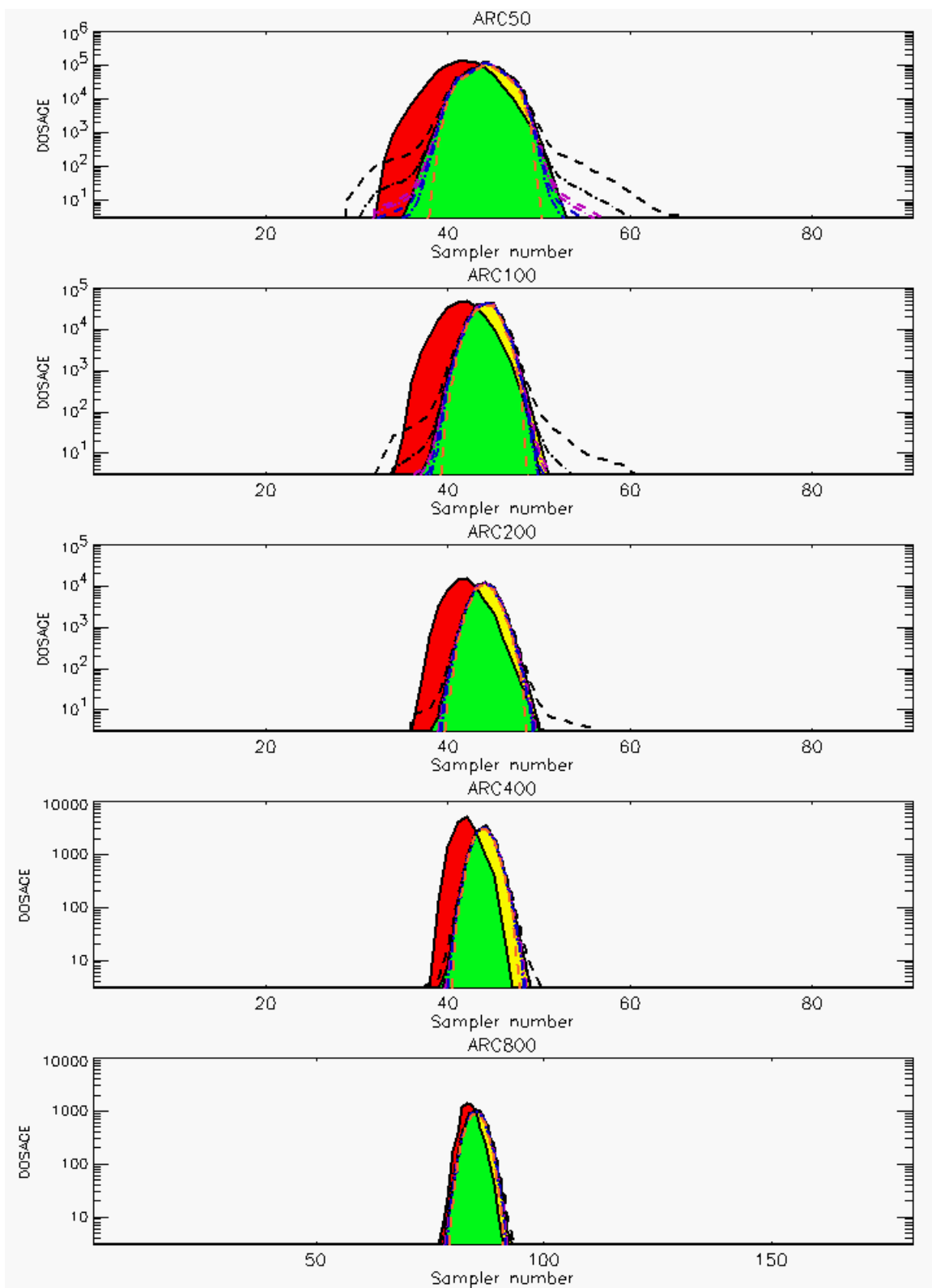


Figure C-17a. HPAC Probabilistic Prediction Outputs for Trial 22 on Linear Scale: Stability Class is 4



**Figure C-17b. HPAC Probabilistic Prediction Outputs for Trial 22 on Logarithmic Scale:
Stability Class is 4**

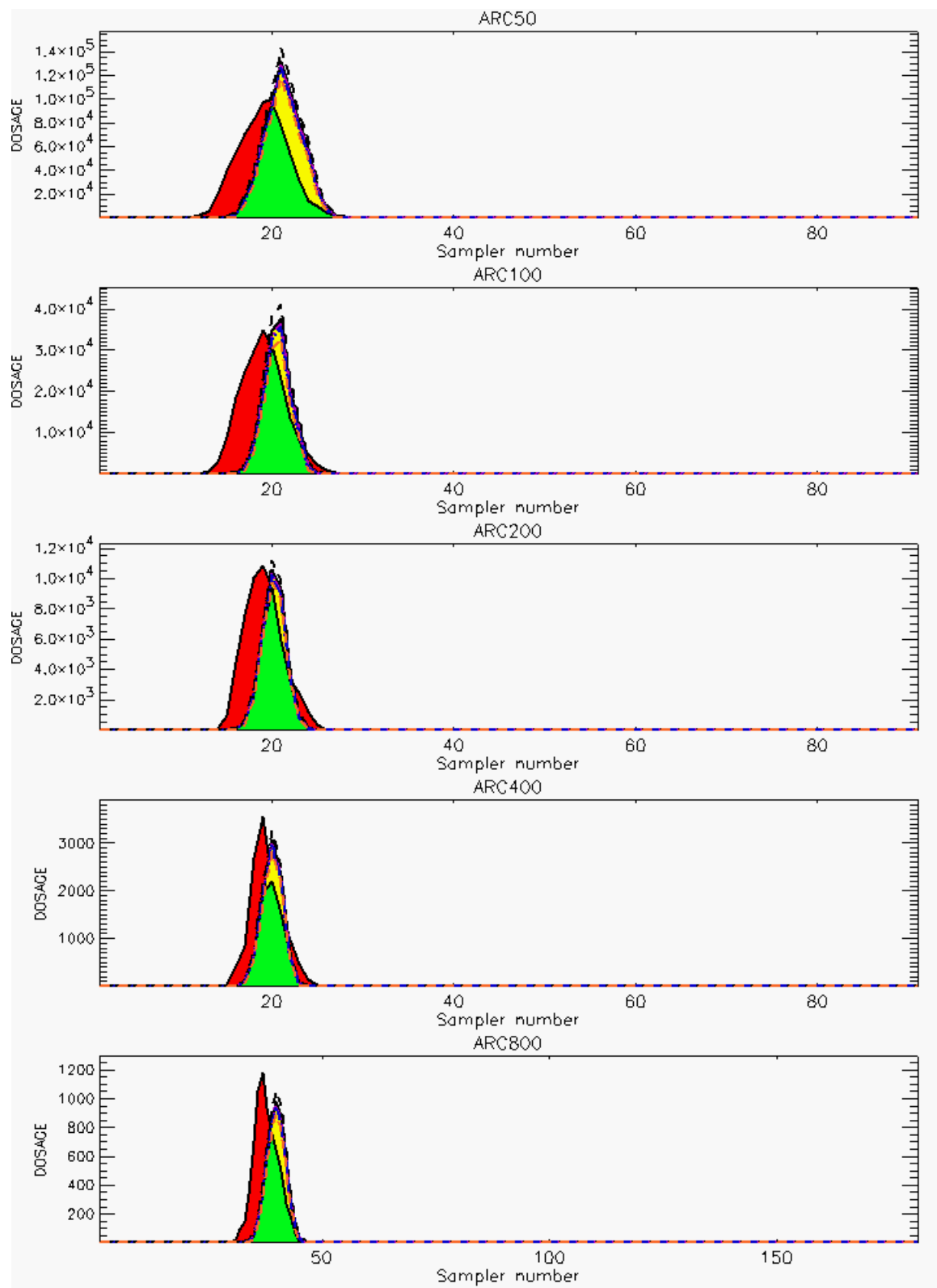
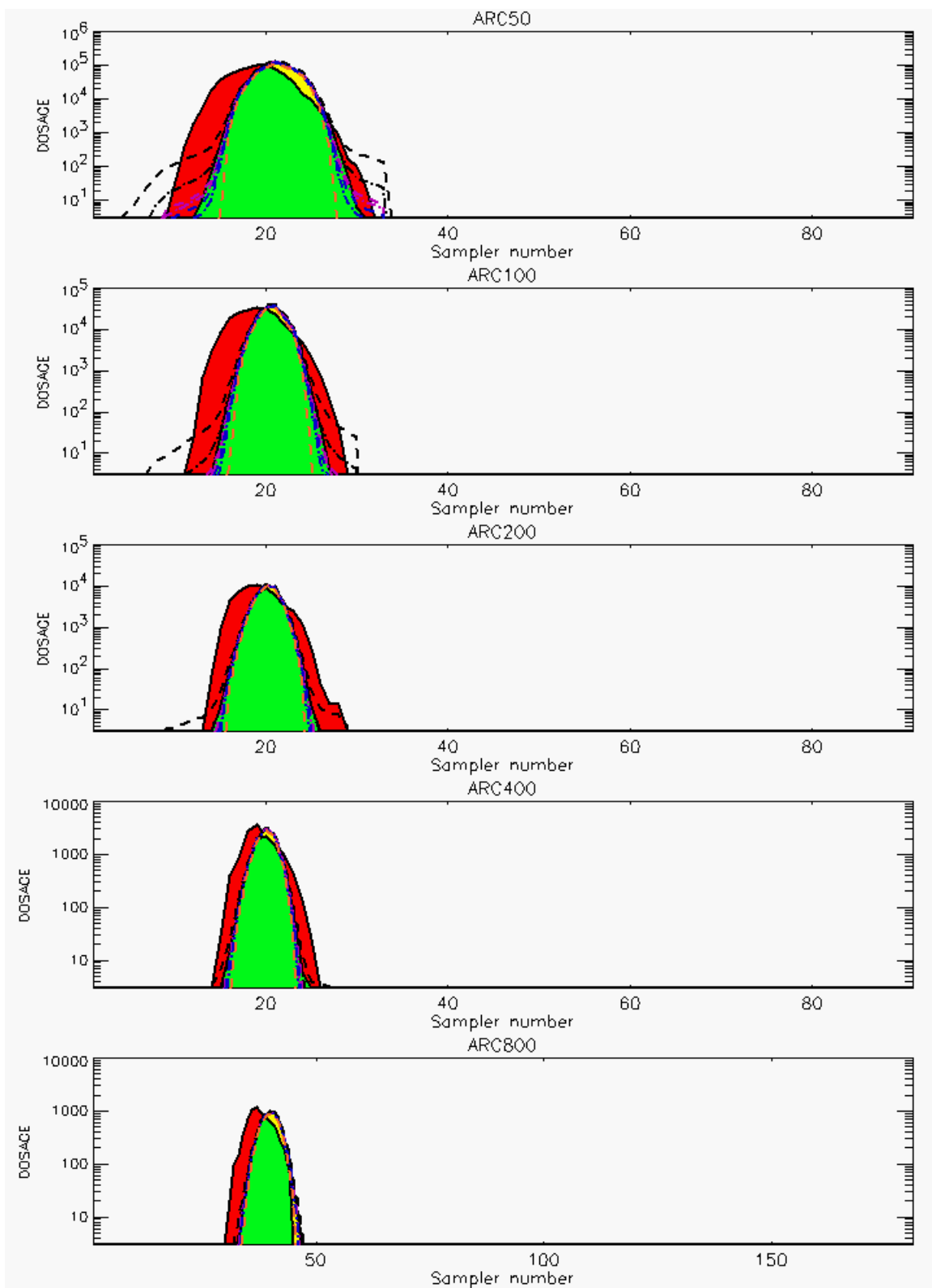


Figure C-18a. HPAC Probabilistic Prediction Outputs for Trial 23 on Linear Scale: Stability Class is 4



**Figure C-18b. HPAC Probabilistic Prediction Outputs for Trial 23 on Logarithmic Scale:
Stability Class is 4**

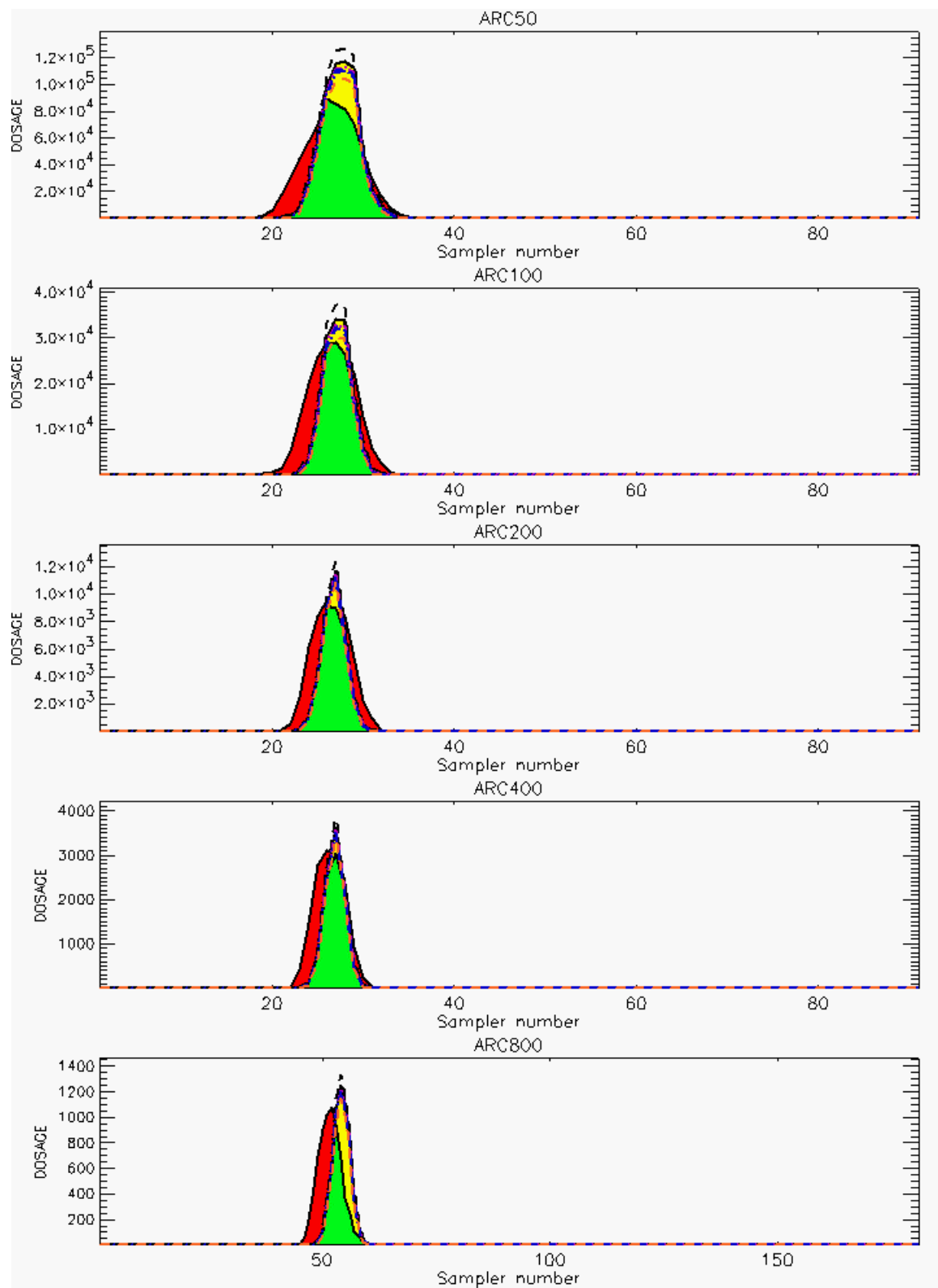
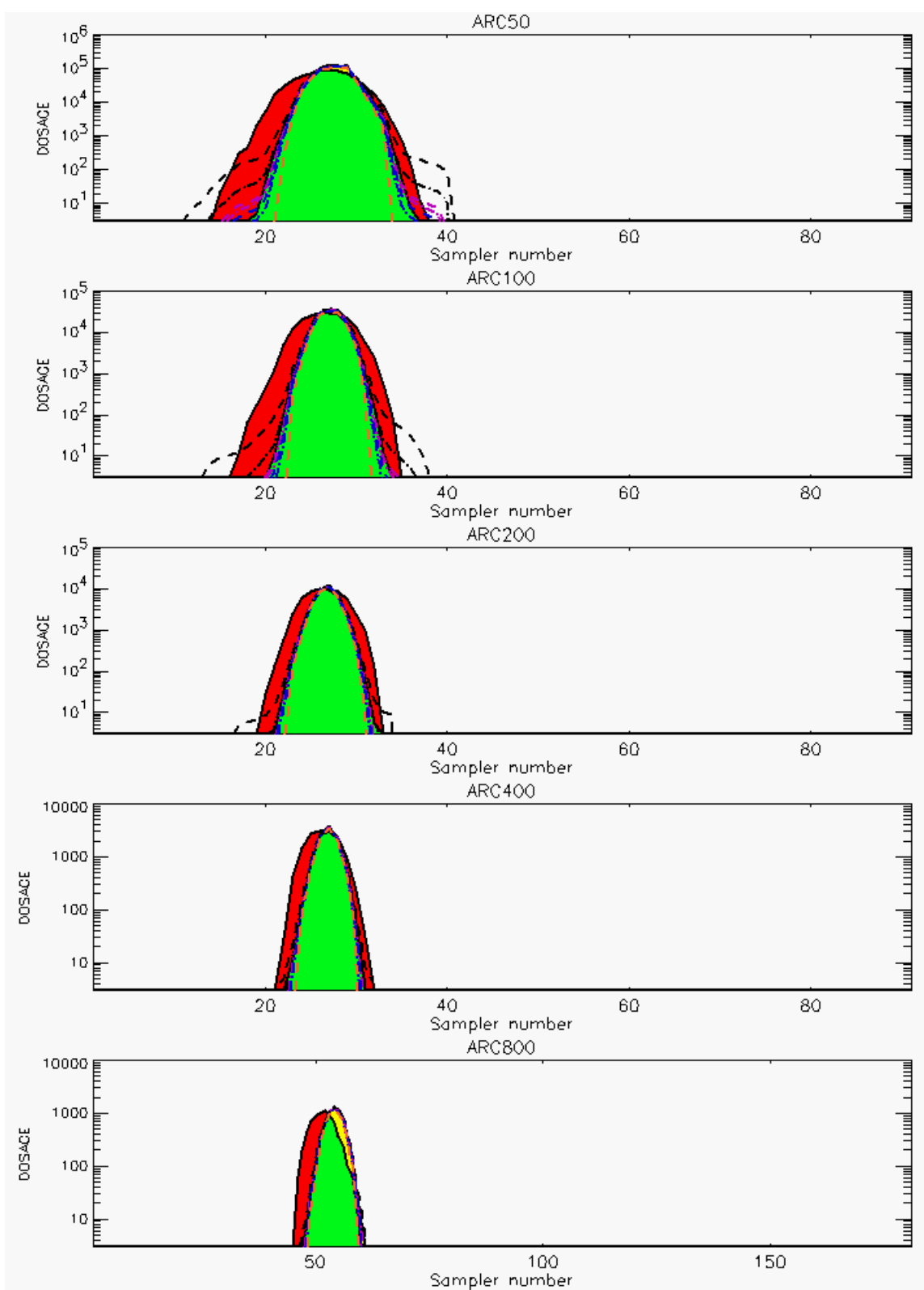


Figure C-19a. HPAC Probabilistic Prediction Outputs for Trial 24 on Linear Scale: Stability Class is 4



**Figure C-19b. HPAC Probabilistic Prediction Outputs for Trial 24 on Logarithmic Scale:
Stability Class is 4**

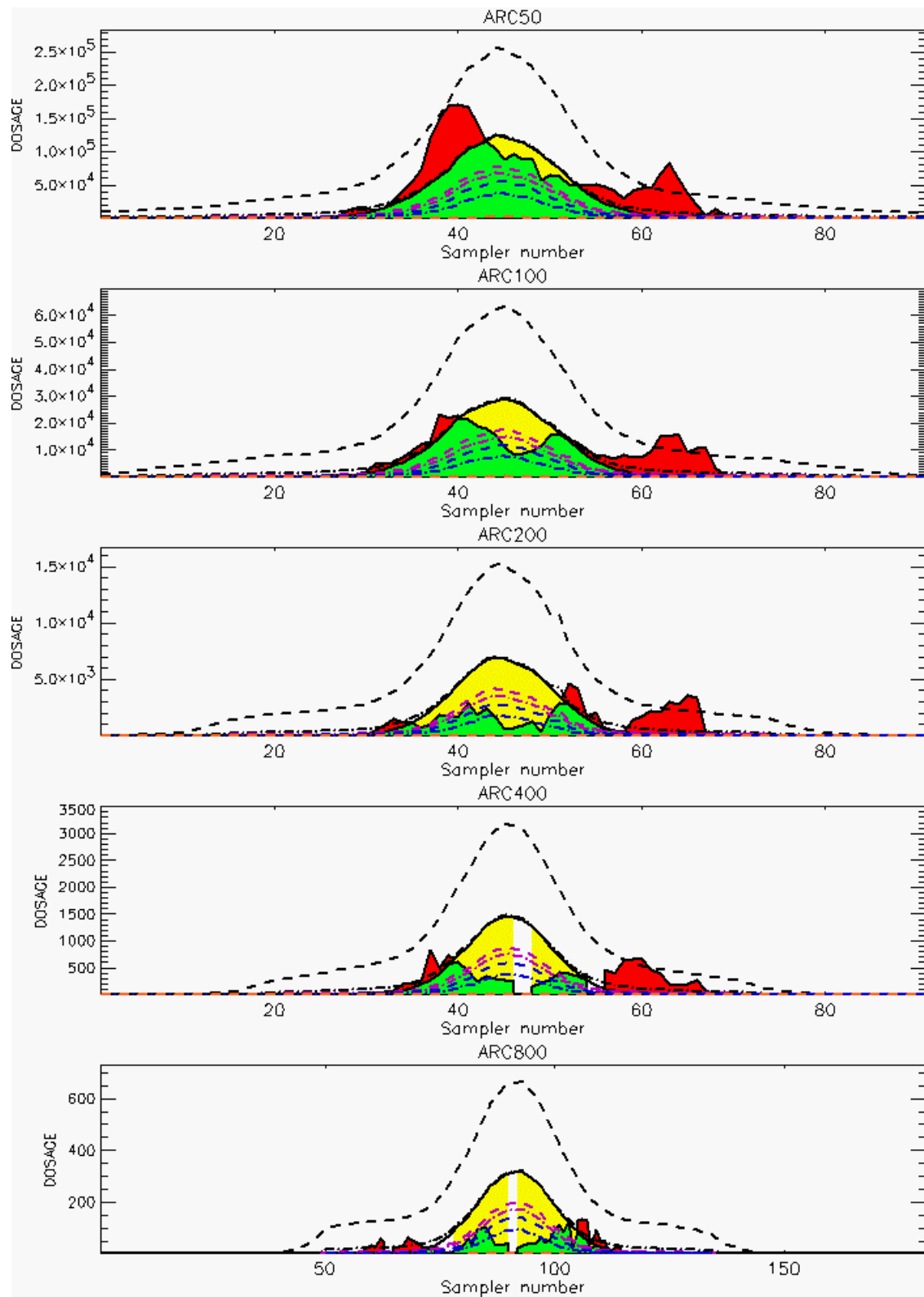


Figure C-20a. HPAC Probabilistic Prediction Outputs for Trial 25 on Linear Scale: Stability Class is 1 (Values for Samplers 47 55 of 400-Meter Arc and 64, 75, 77, 91 of 800-Meter Arc are Missing)

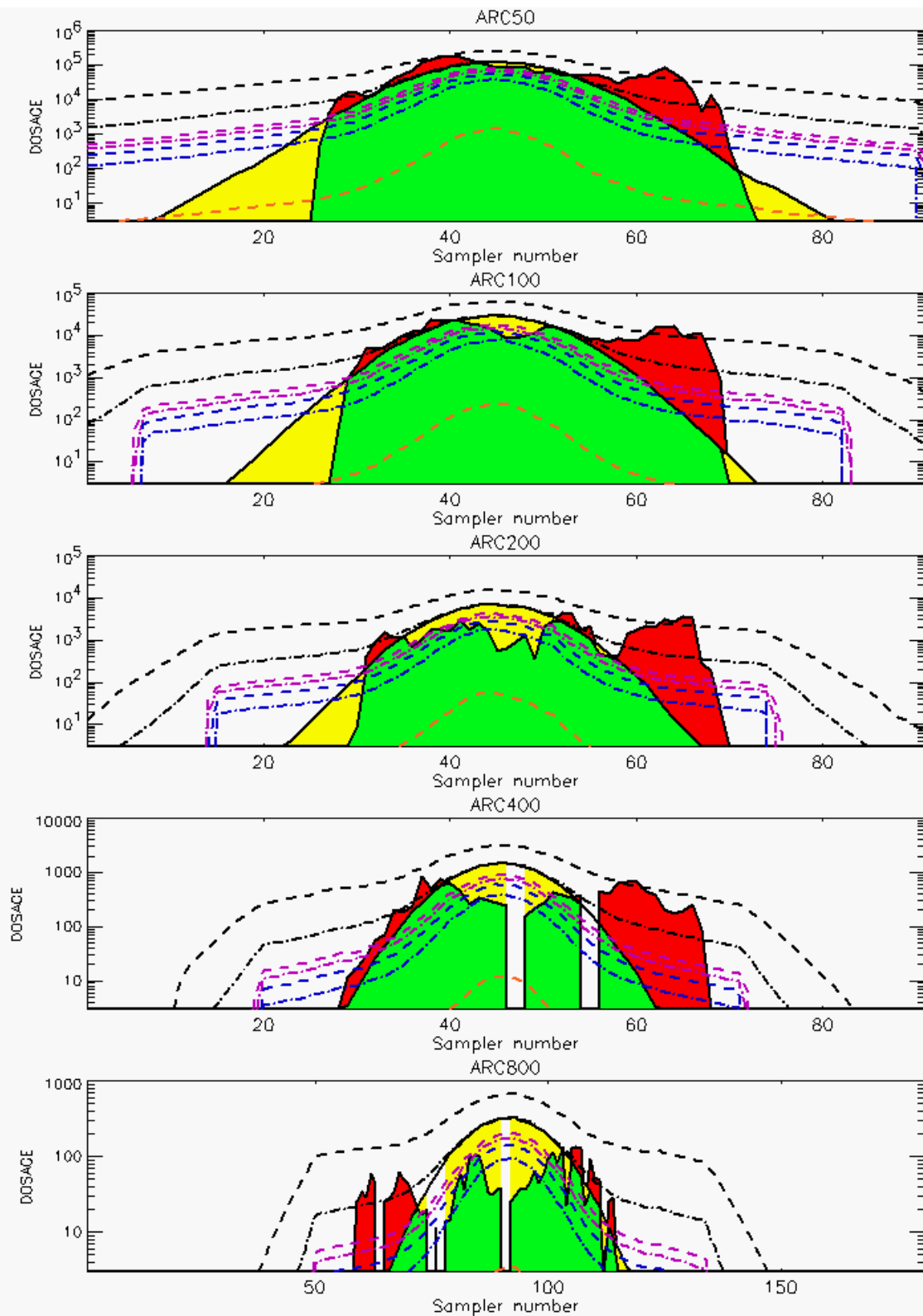


Figure C-20b. HPAC Probabilistic Prediction Outputs for Trial 25 on Logarithmic Scale: Stability Class is 1 (Values for Samplers 47 55 of 400-Meter Arc and 64, 75, 77, 91 of 800-Meter Arc are Missing)

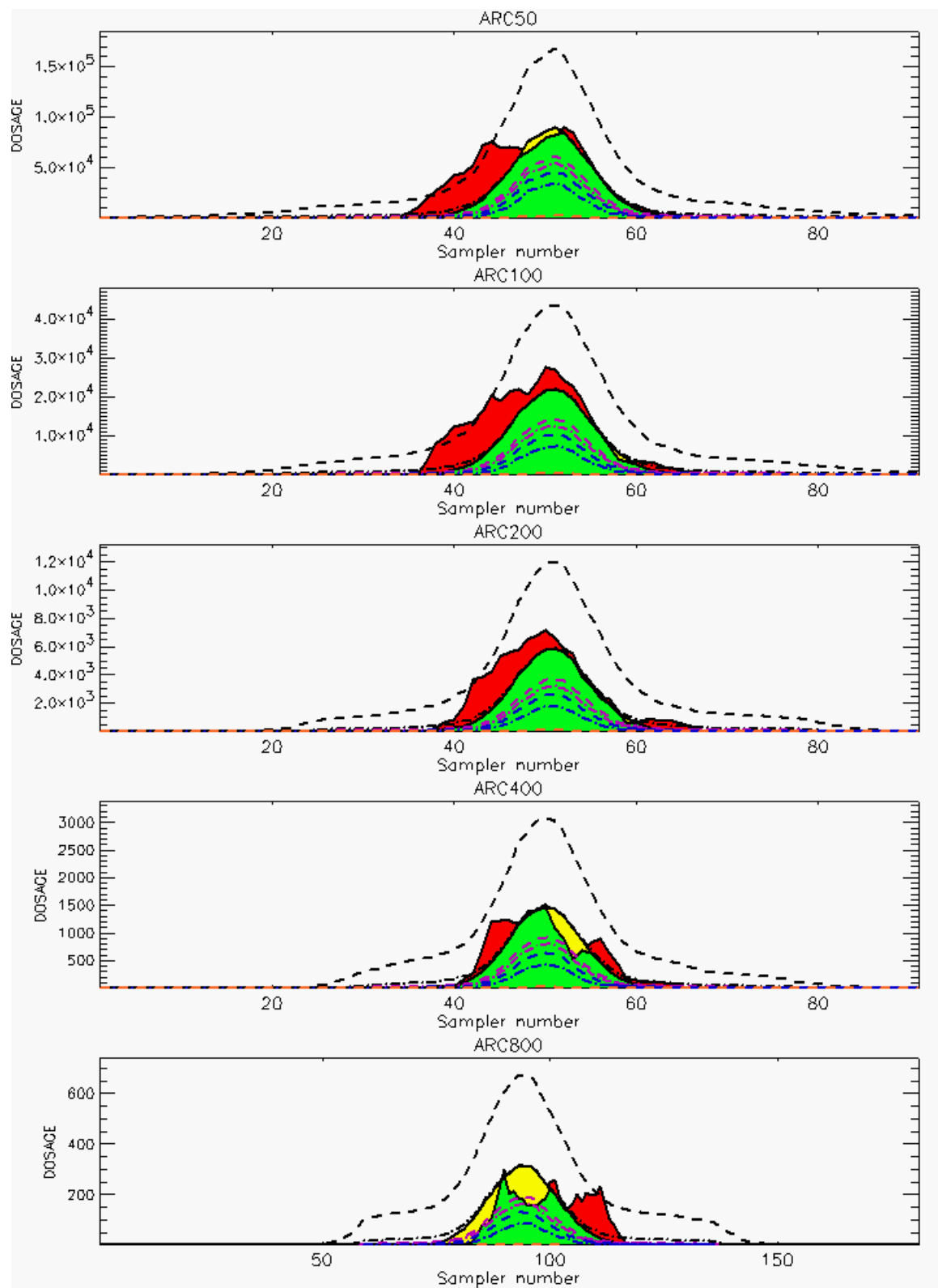


Figure C-21a. HPAC Probabilistic Prediction Outputs for Trial 26 on Linear Scale: Stability Class is 2

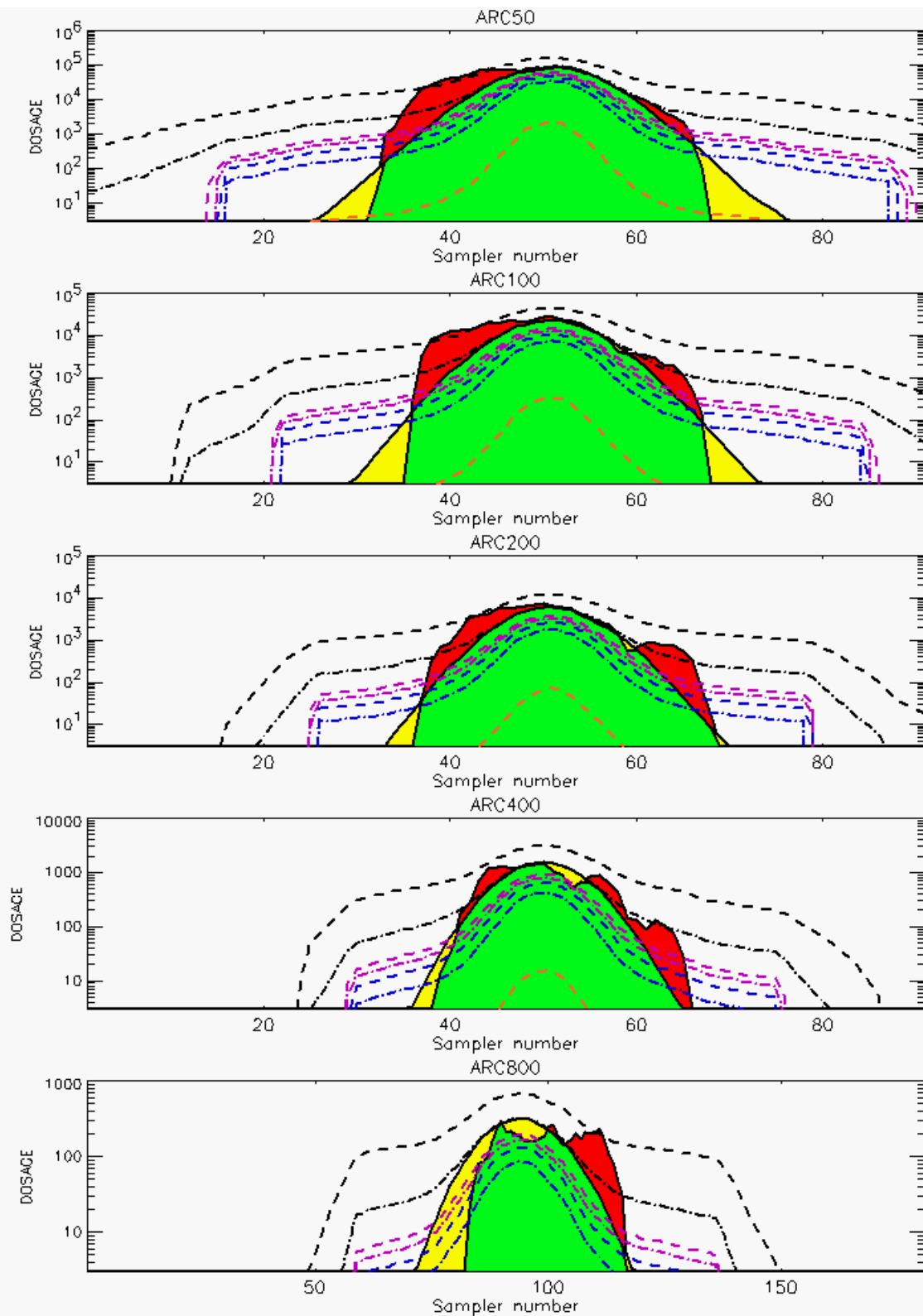


Figure C-21b. HPAC Probabilistic Prediction Outputs for Trial 26 on Logarithmic Scale: Stability Class is 2

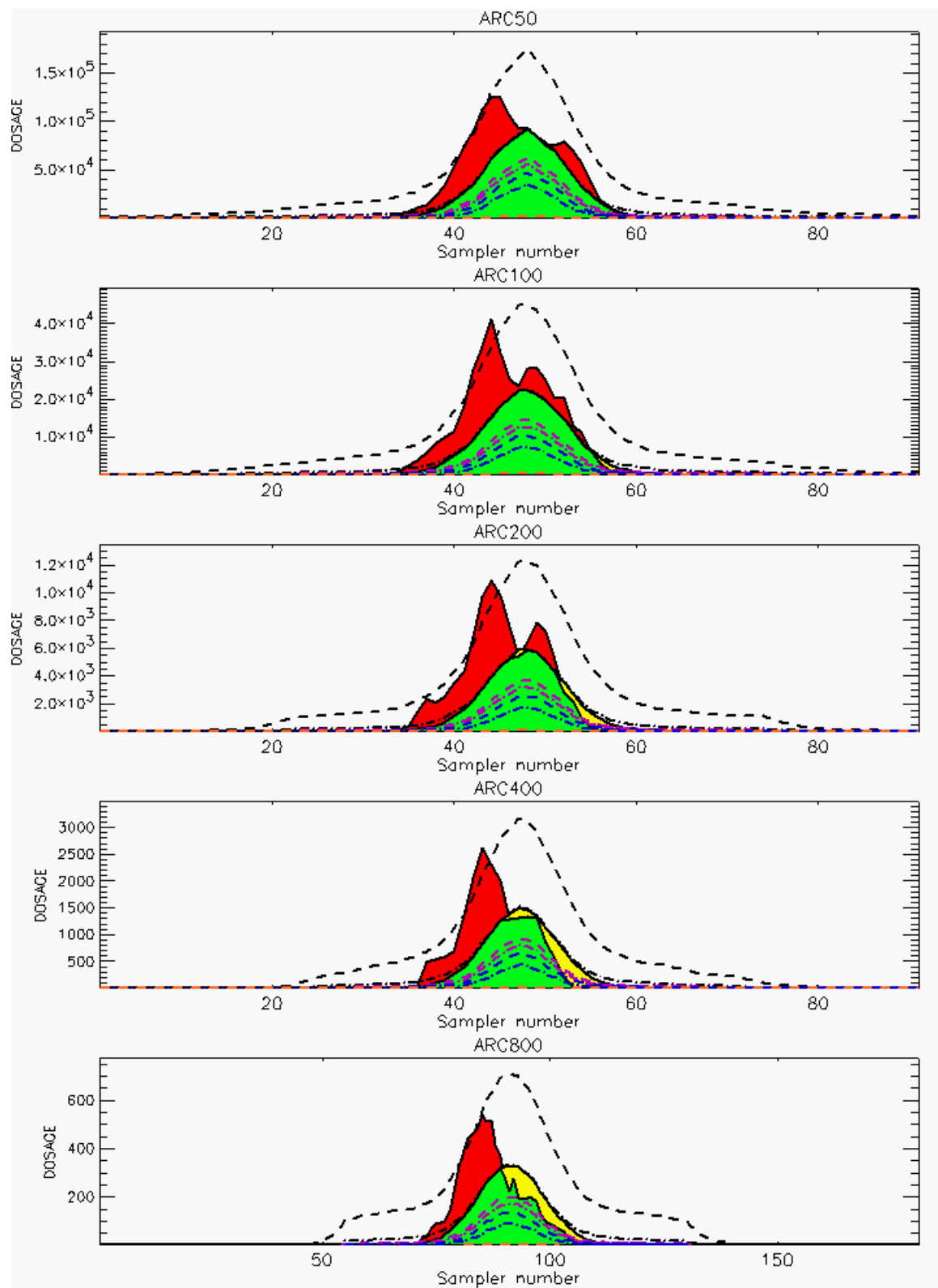


Figure C-22a. HPAC Probabilistic Prediction Outputs for Trial 27 on Linear Scale: Stability Class is 2

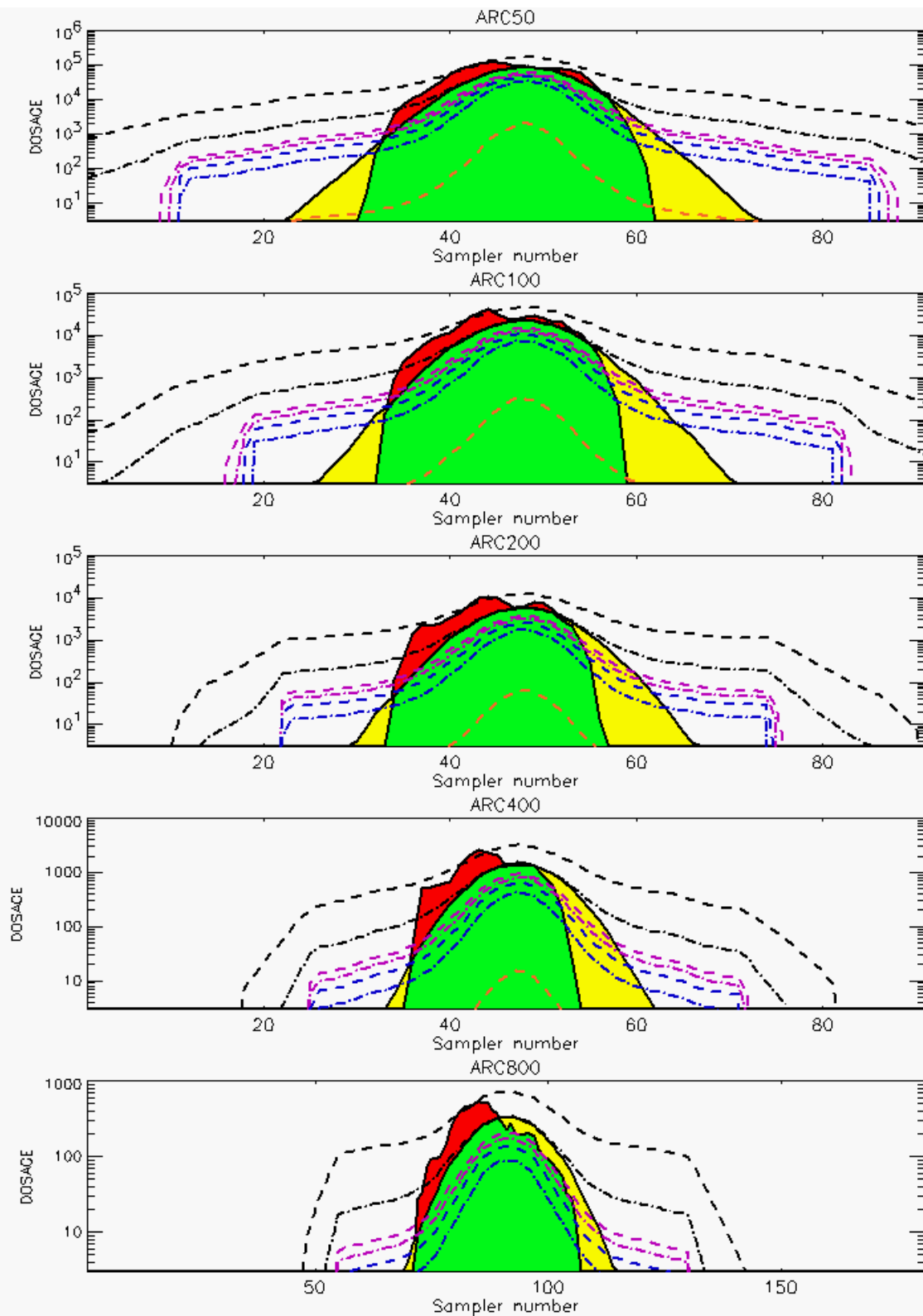


Figure C-22b. HPAC Probabilistic Prediction Outputs for Trial 27 on Logarithmic Scale: Stability Class is 2

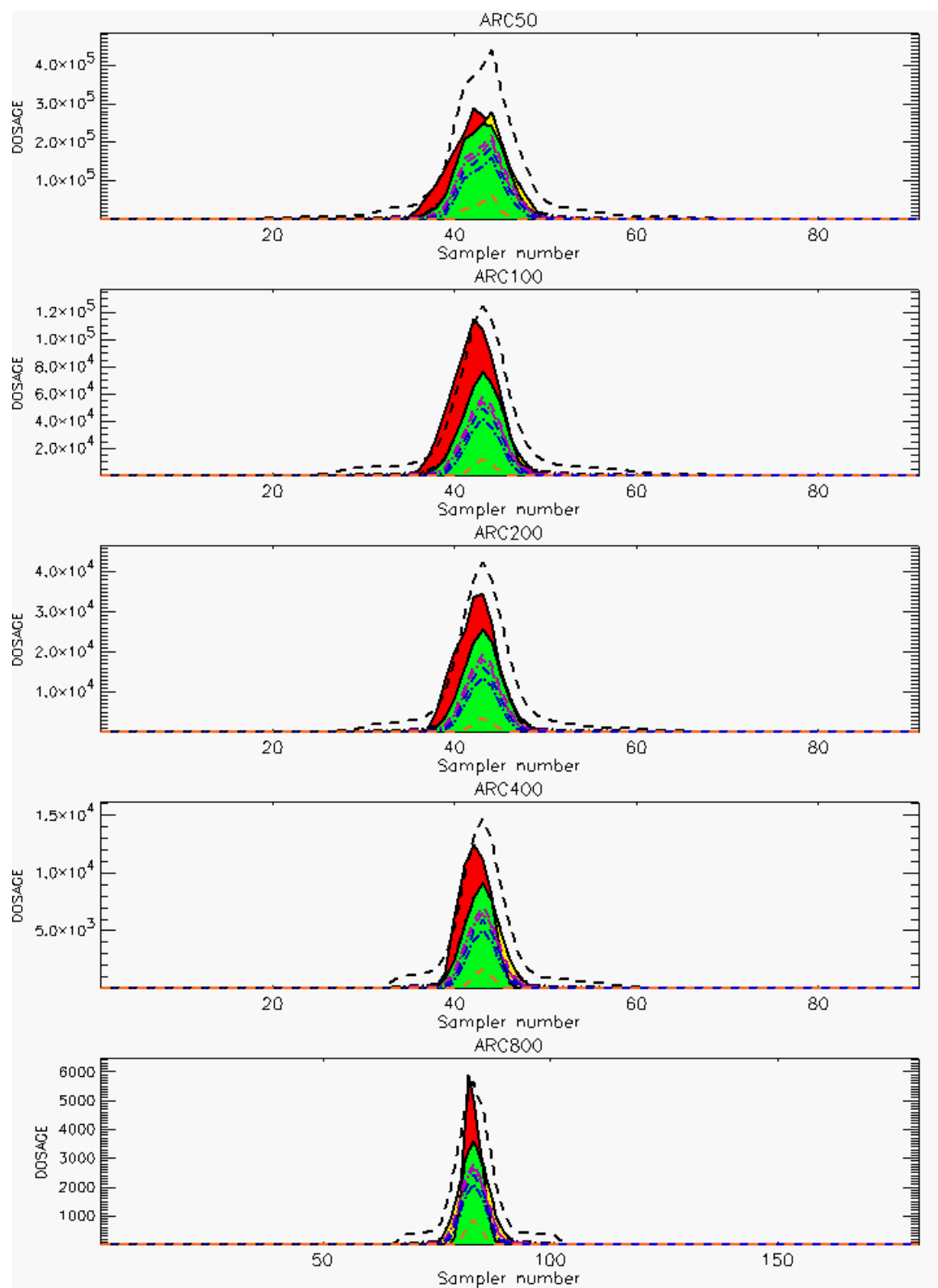
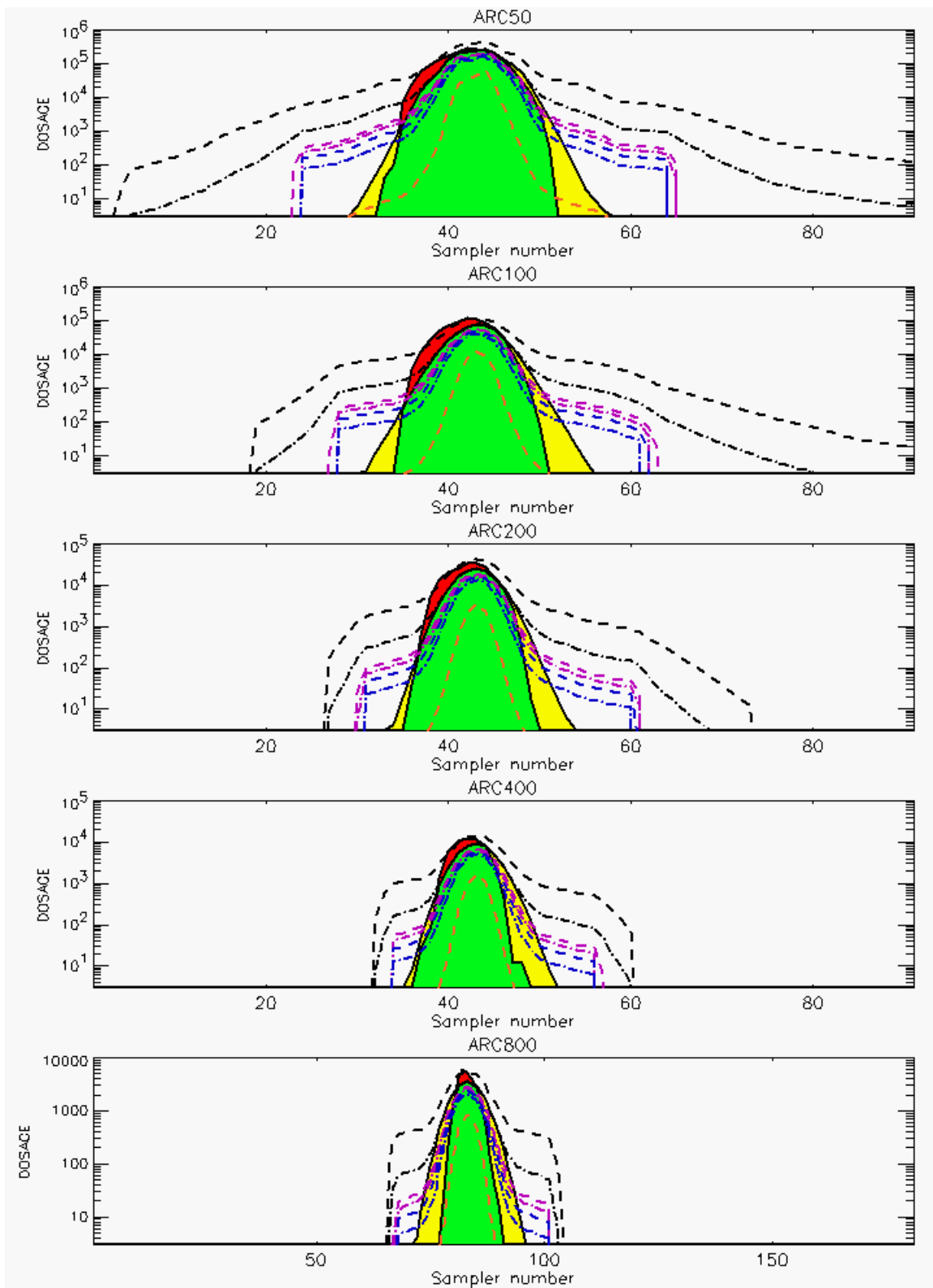


Figure C-23a. HPAC Probabilistic Prediction Outputs for Trial 28 on Linear Scale: Stability Class is 5



**Figure C-23b. HPAC Probabilistic Prediction Outputs for Trial 28 on Logarithmic Scale:
Stability Class is 5**

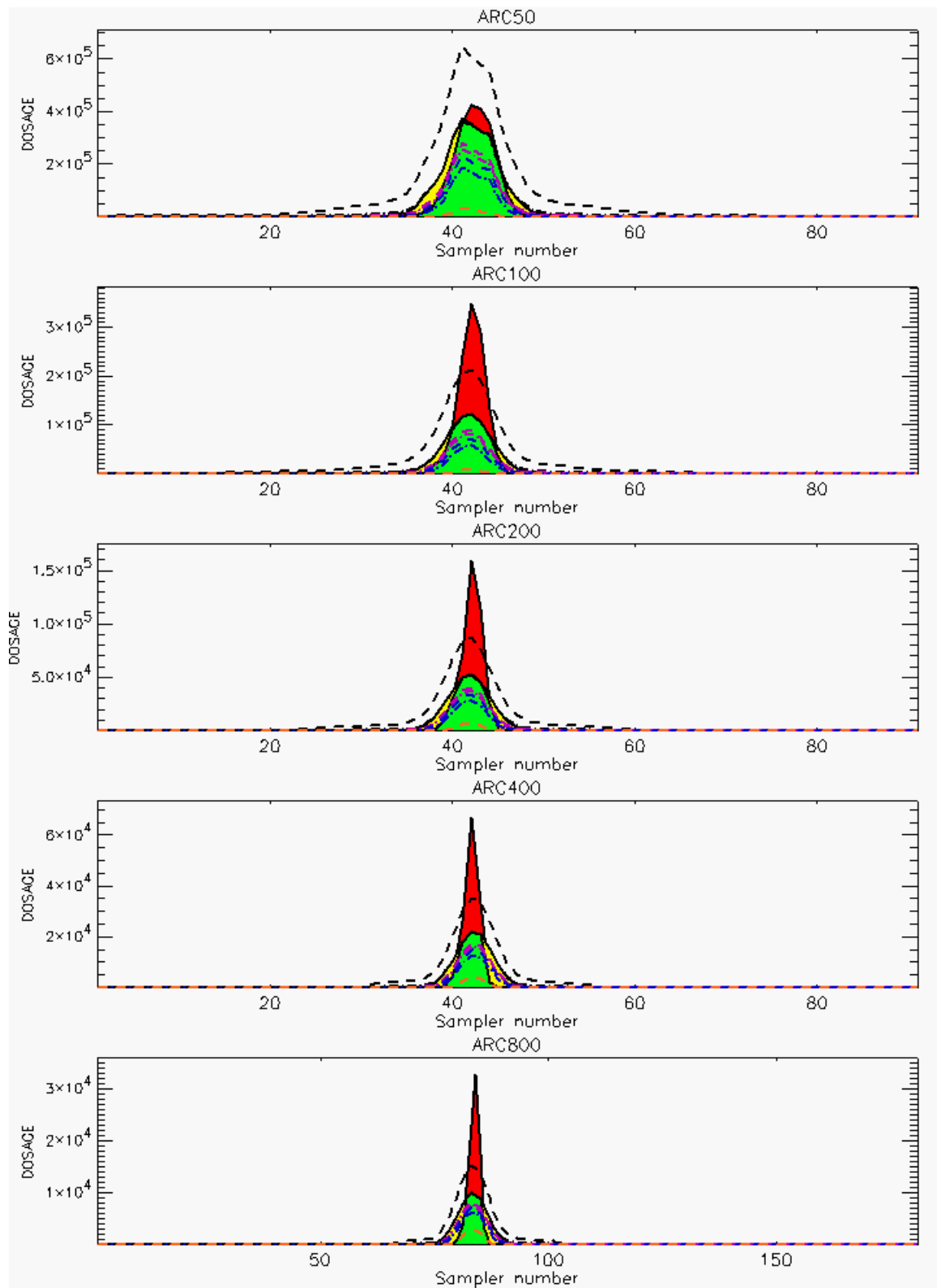


Figure C-24a. HPAC Probabilistic Prediction Outputs for Trial 32 on Linear Scale: Stability Class is 6

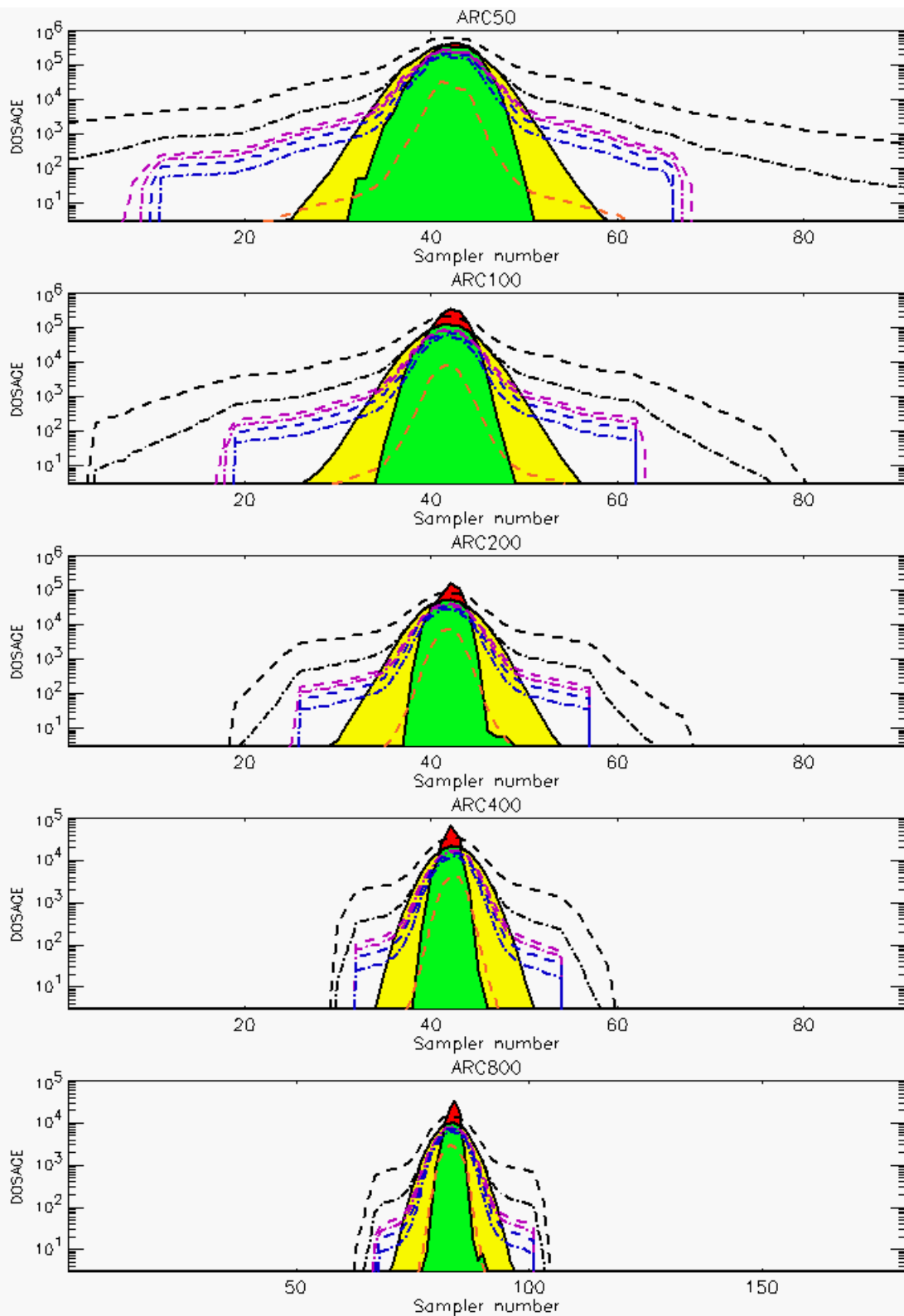


Figure C-24b. HPAC Probabilistic Prediction Outputs for Trial 32 on Logarithmic Scale: Stability Class is 6

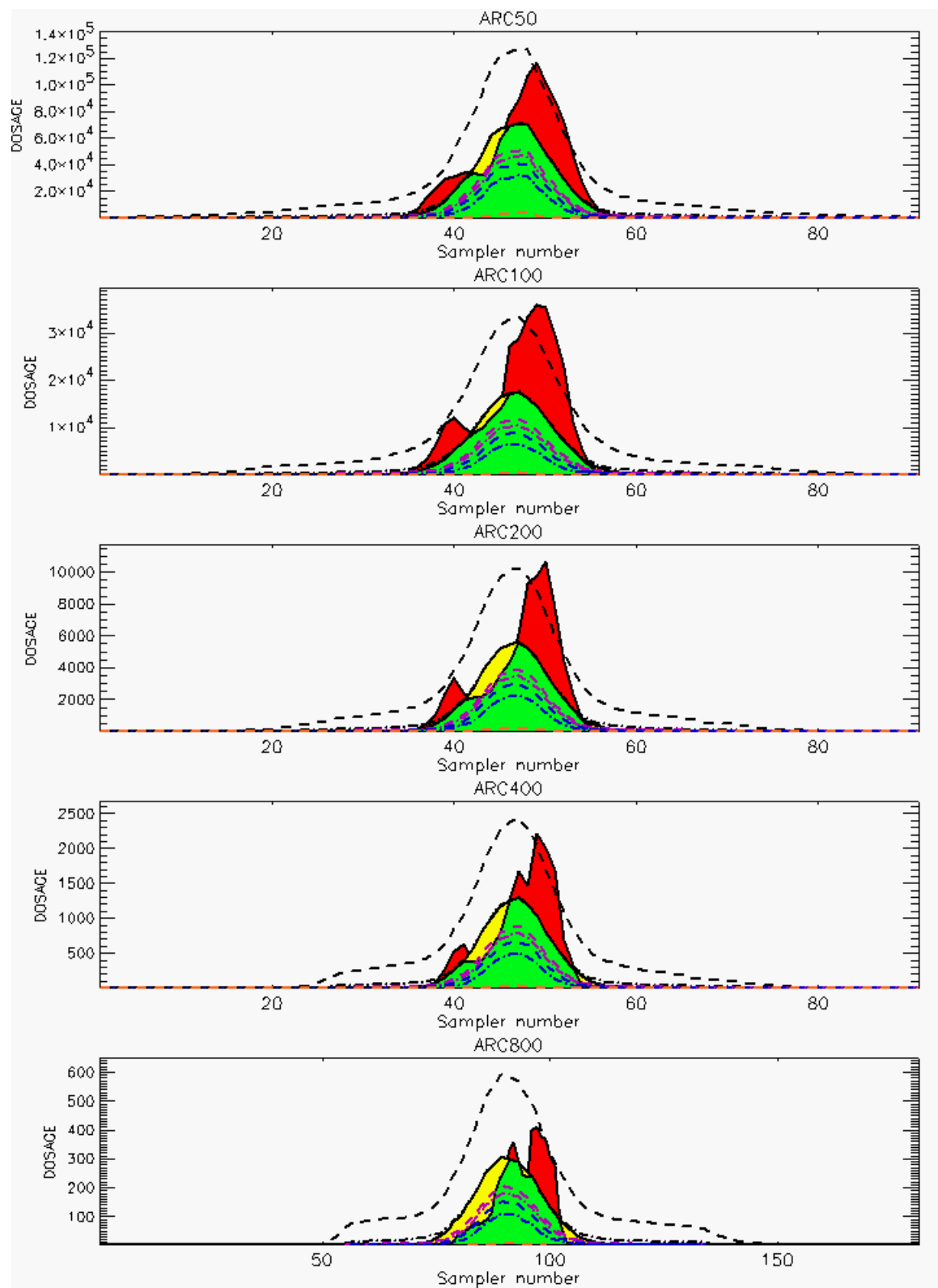


Figure C-25a. HPAC Probabilistic Prediction Outputs for Trial 33 on Linear Scale: Stability Class is 3

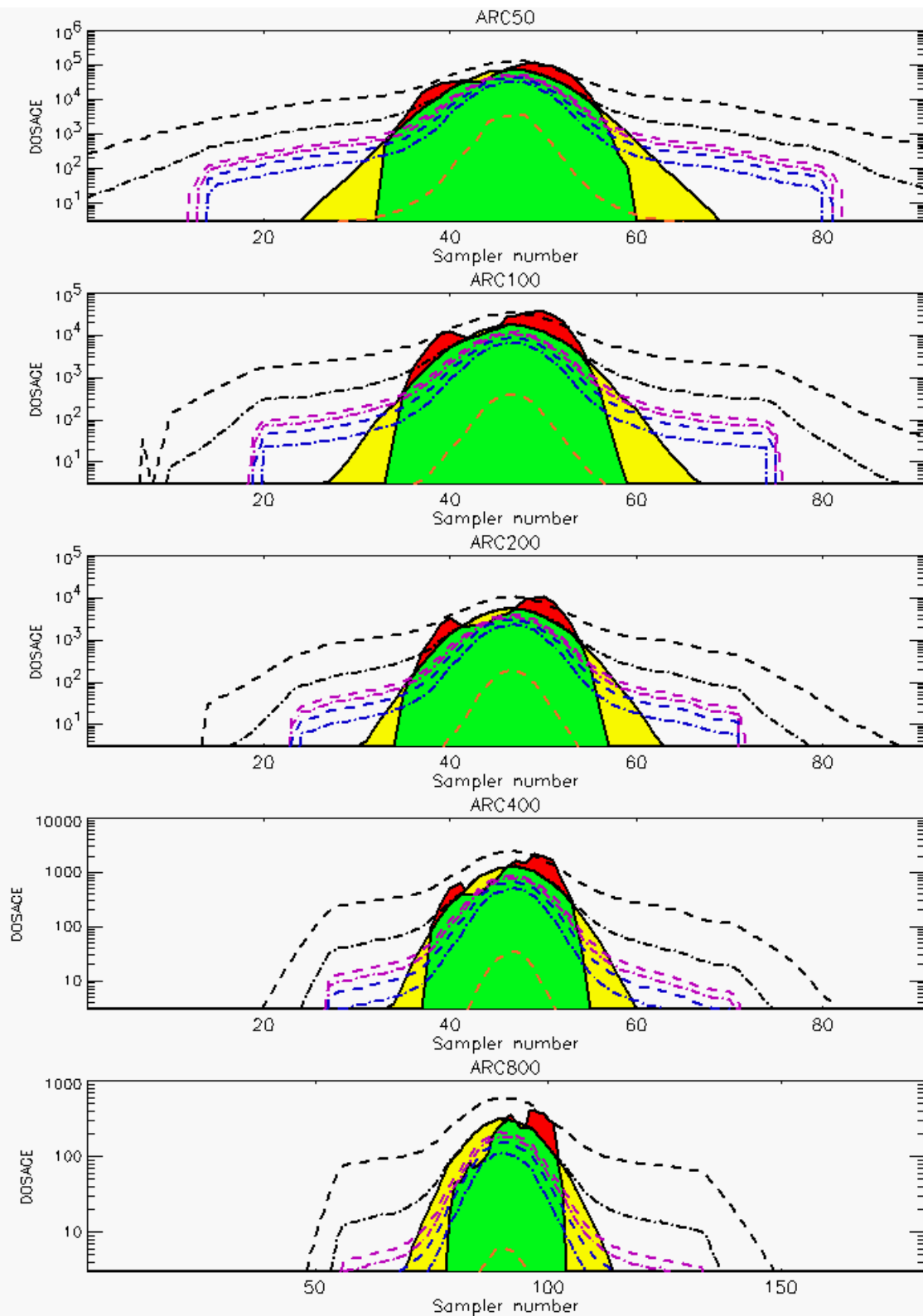


Figure C-25b. HPAC Probabilistic Prediction Outputs for Trial 33 on Logarithmic Scale: Stability Class is 3

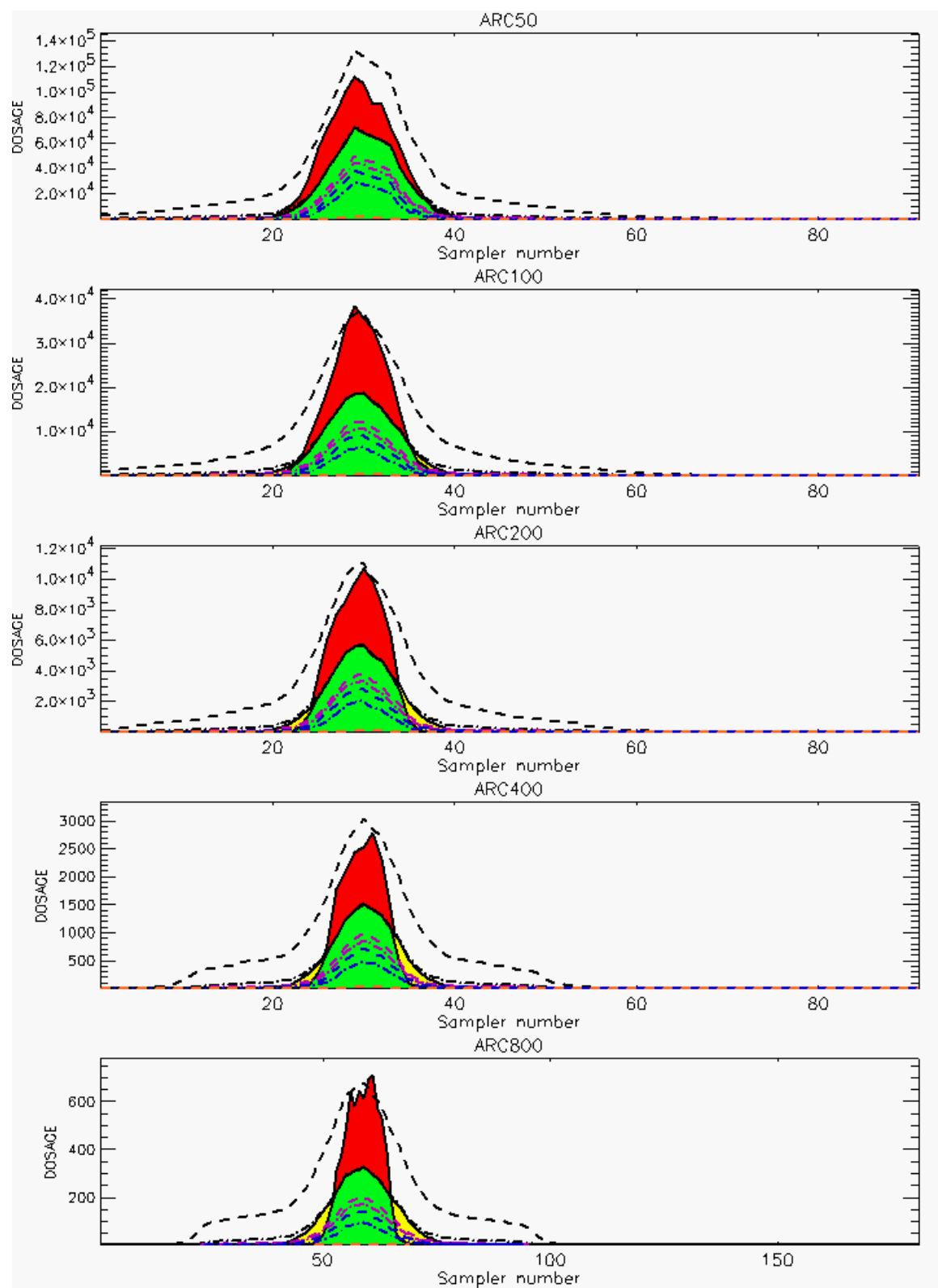


Figure C-26a. HPAC Probabilistic Prediction Outputs for Trial 34 on Linear Scale: Stability Class is 3

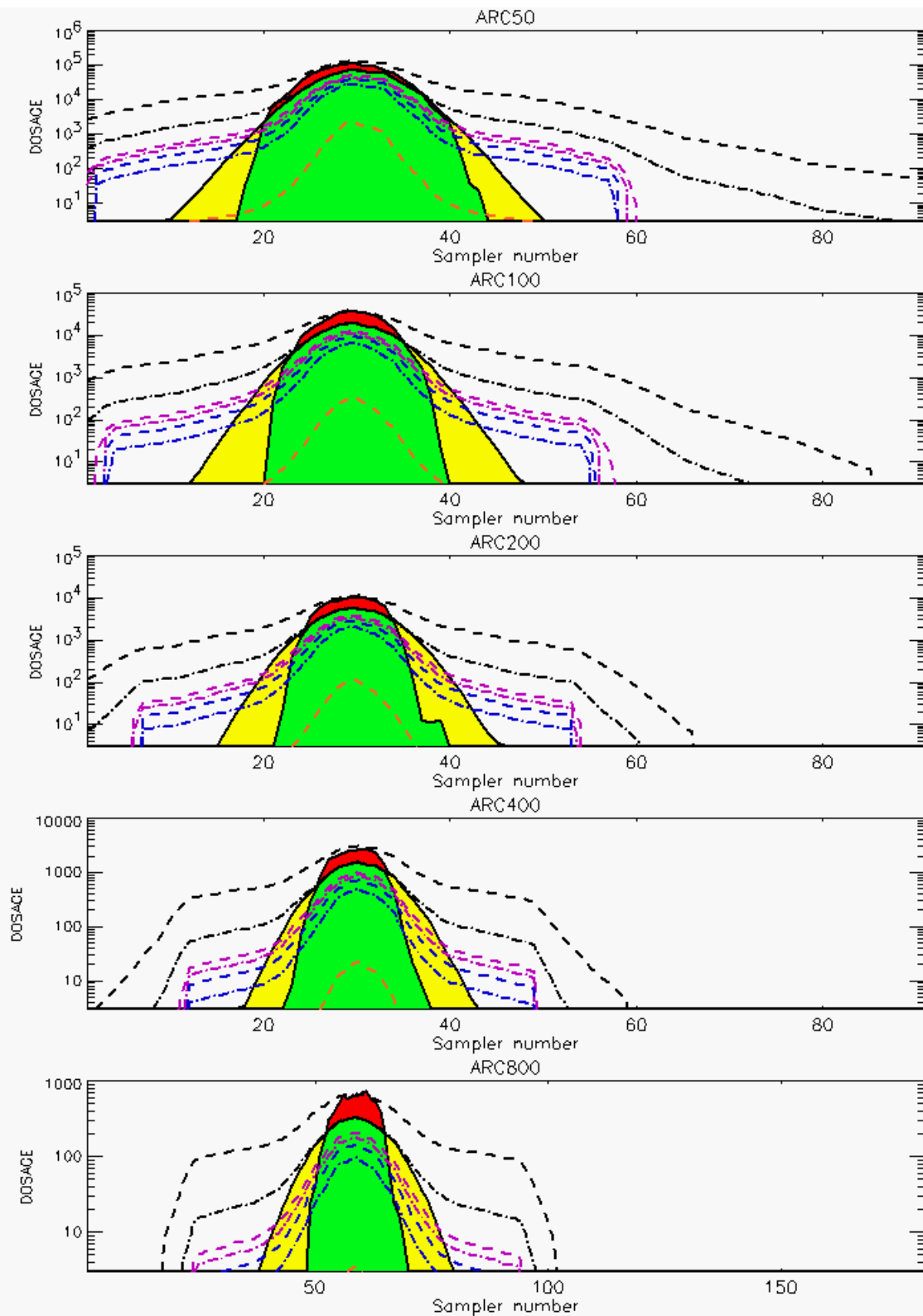


Figure C-26b. HPAC Probabilistic Prediction Outputs for Trial 34 on Logarithmic Scale: Stability Class is 3

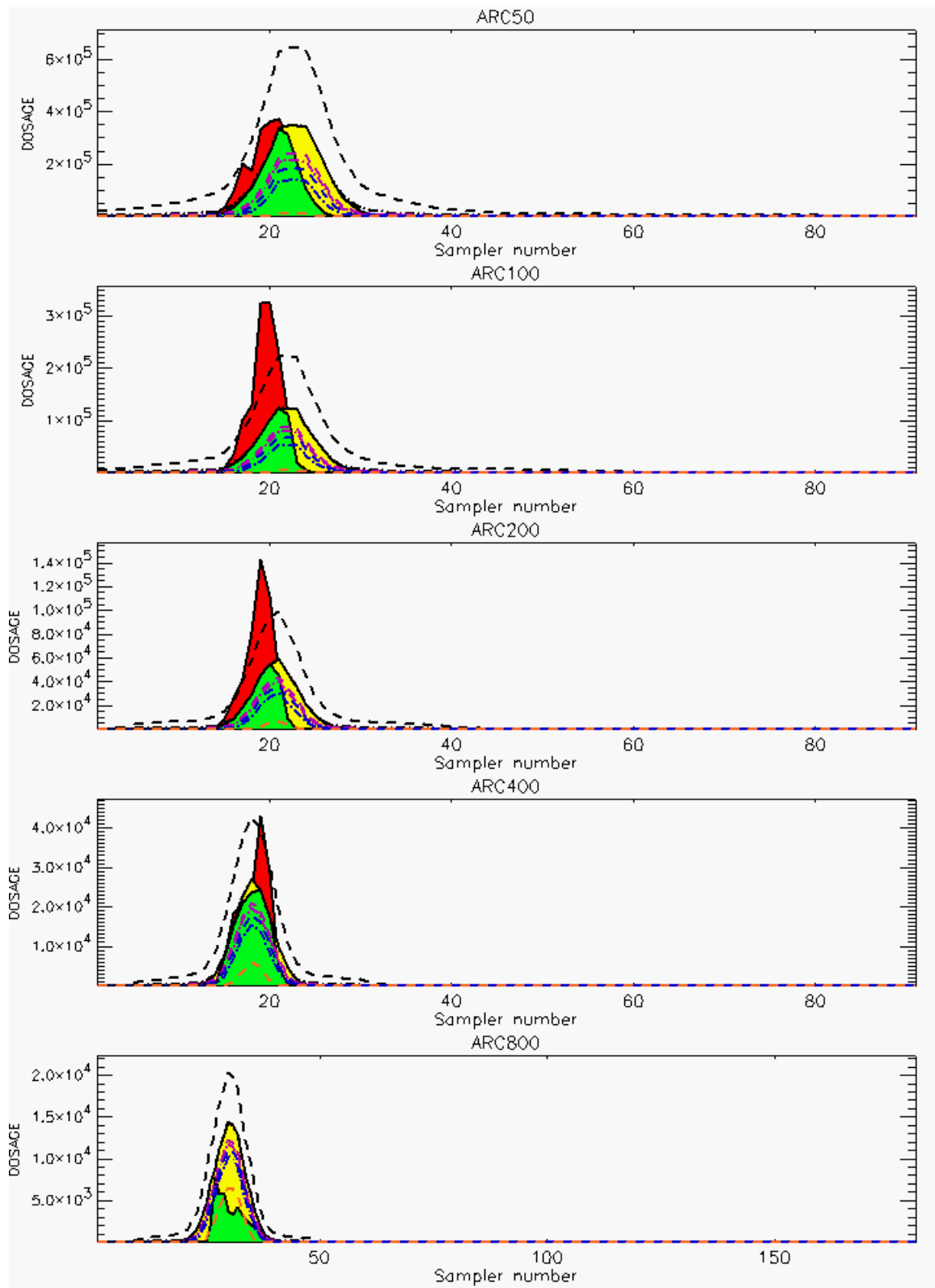
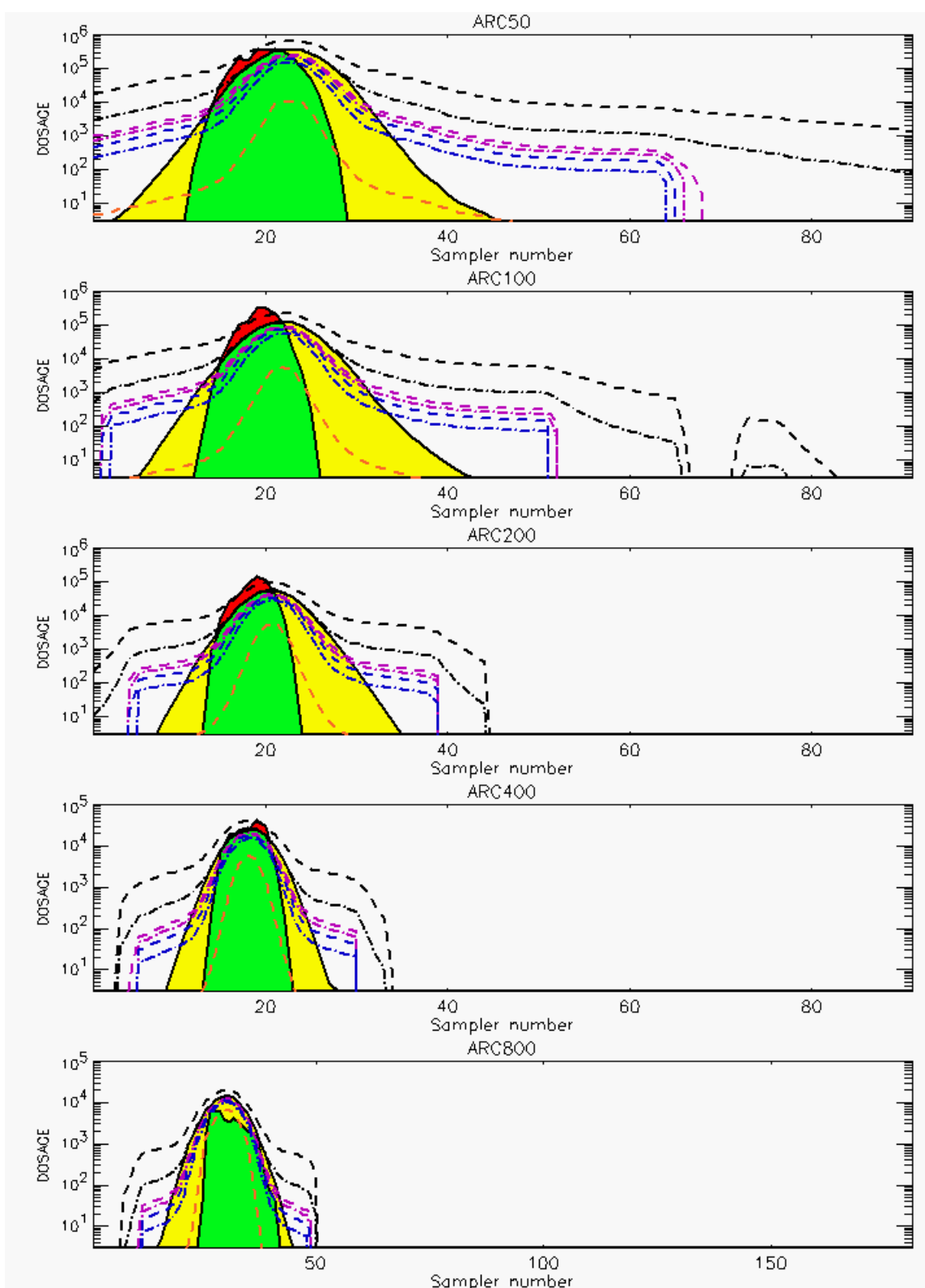


Figure C-27a. HPAC Probabilistic Prediction Outputs for Trial 35 on Linear Scale: Stability Class is 6



**Figure C-27b. HPAC Probabilistic Prediction Outputs for Trial 35 on Logarithmic Scale:
Stability Class is 6**

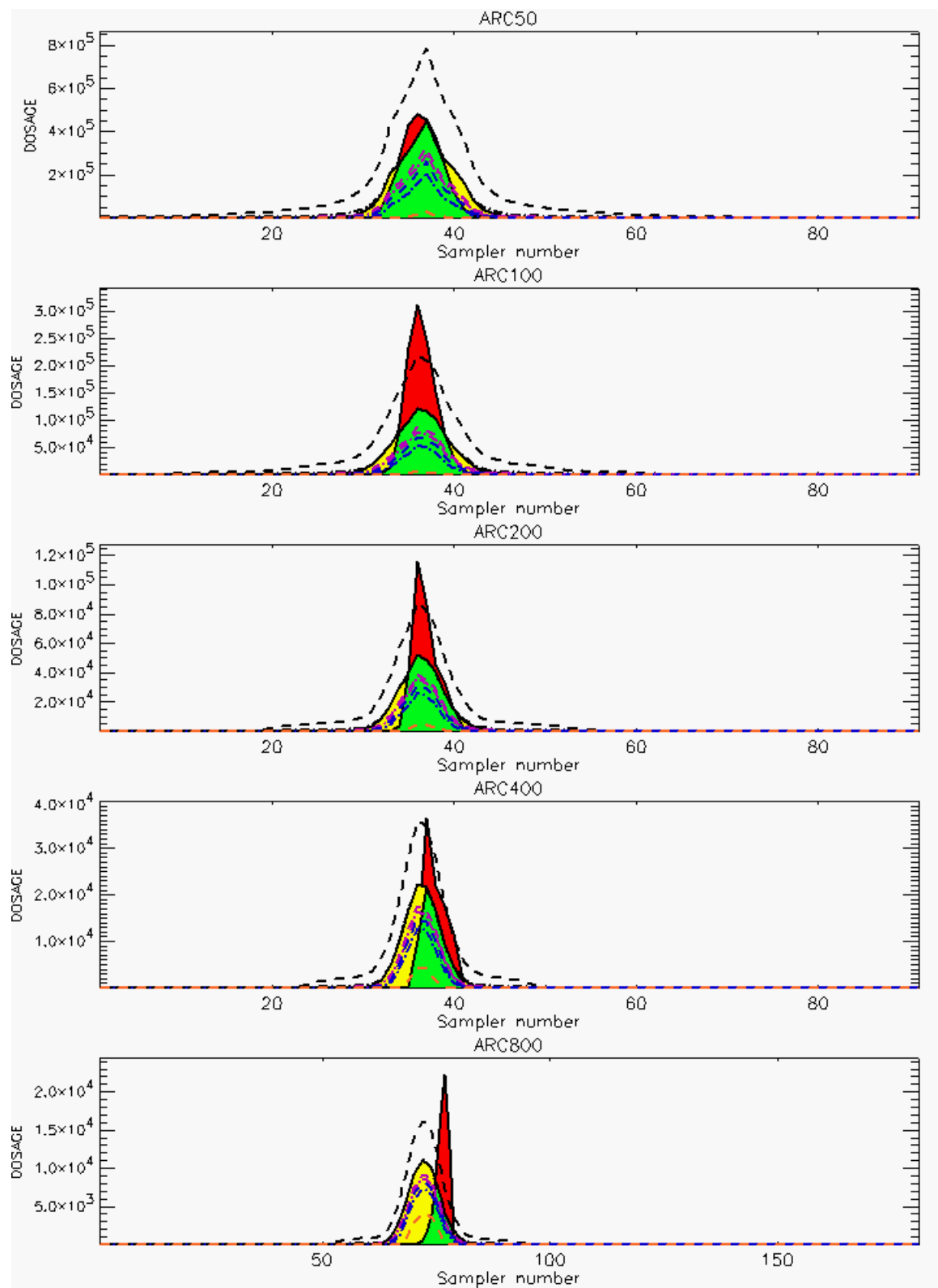
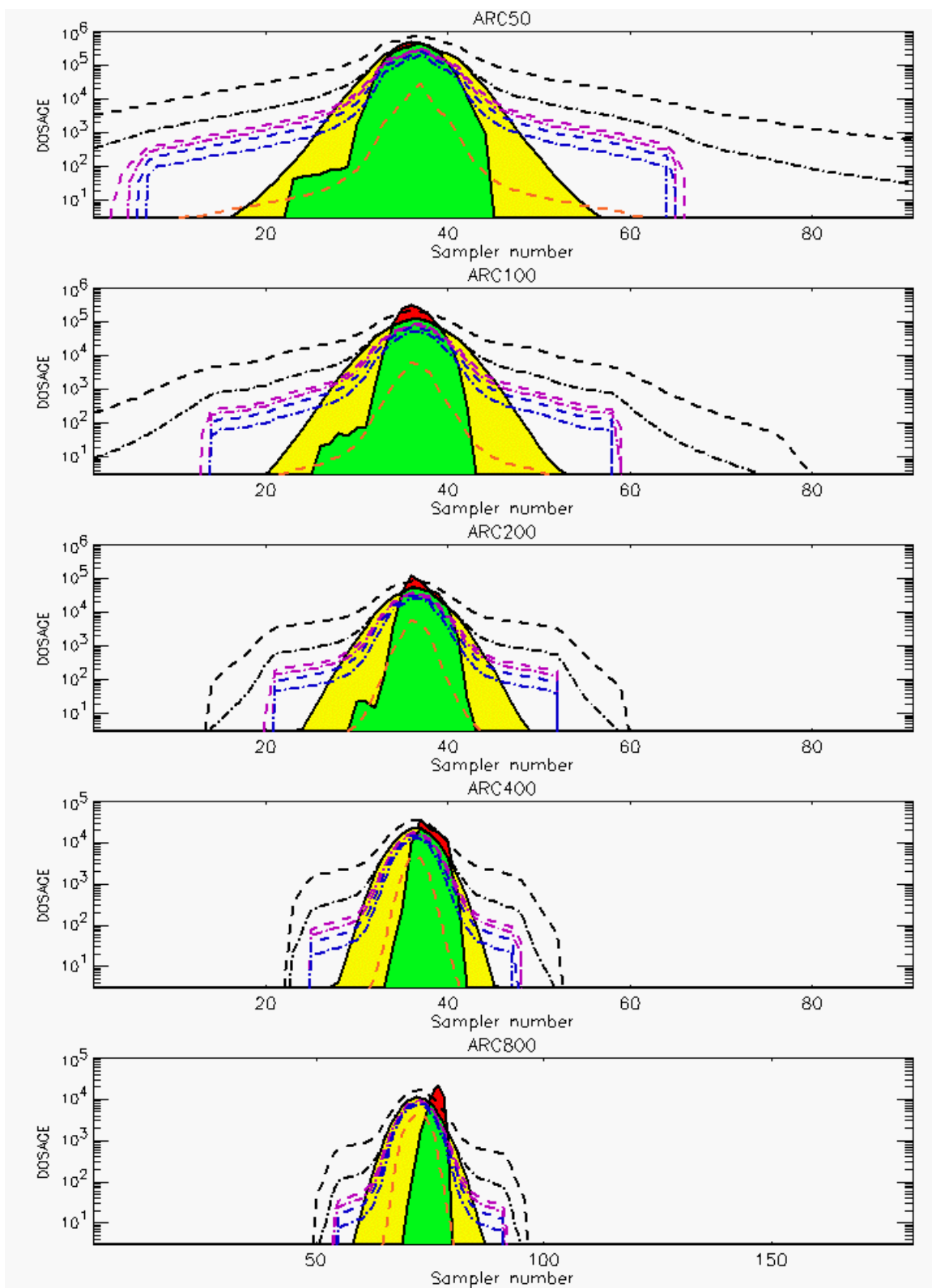


Figure C-28a. HPAC Probabilistic Prediction Outputs for Trial 36 on Linear Scale: Stability Class is 6



**Figure C-28b. HPAC Probabilistic Prediction Outputs for Trial 36 on Logarithmic Scale:
Stability Class is 6**

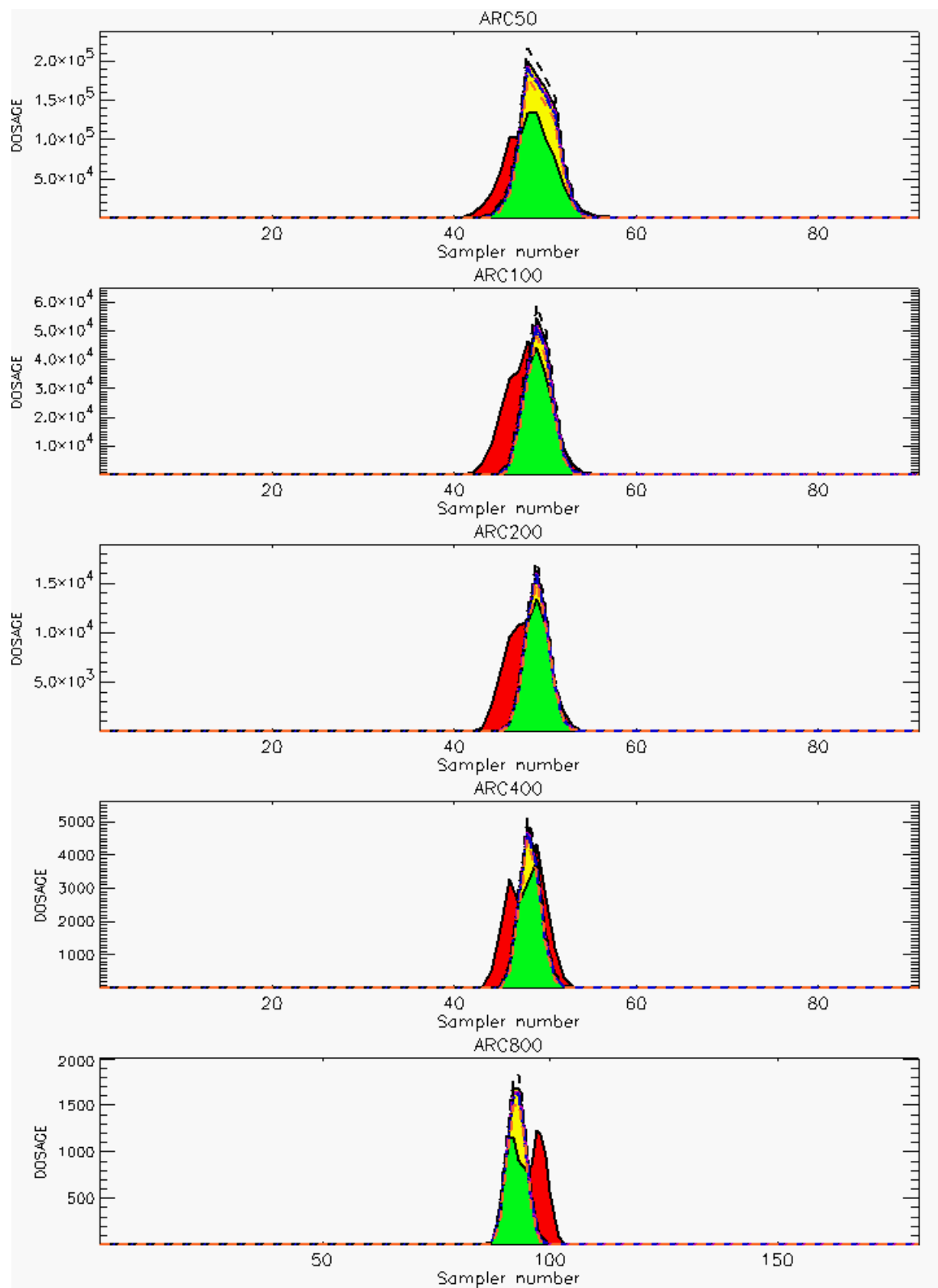
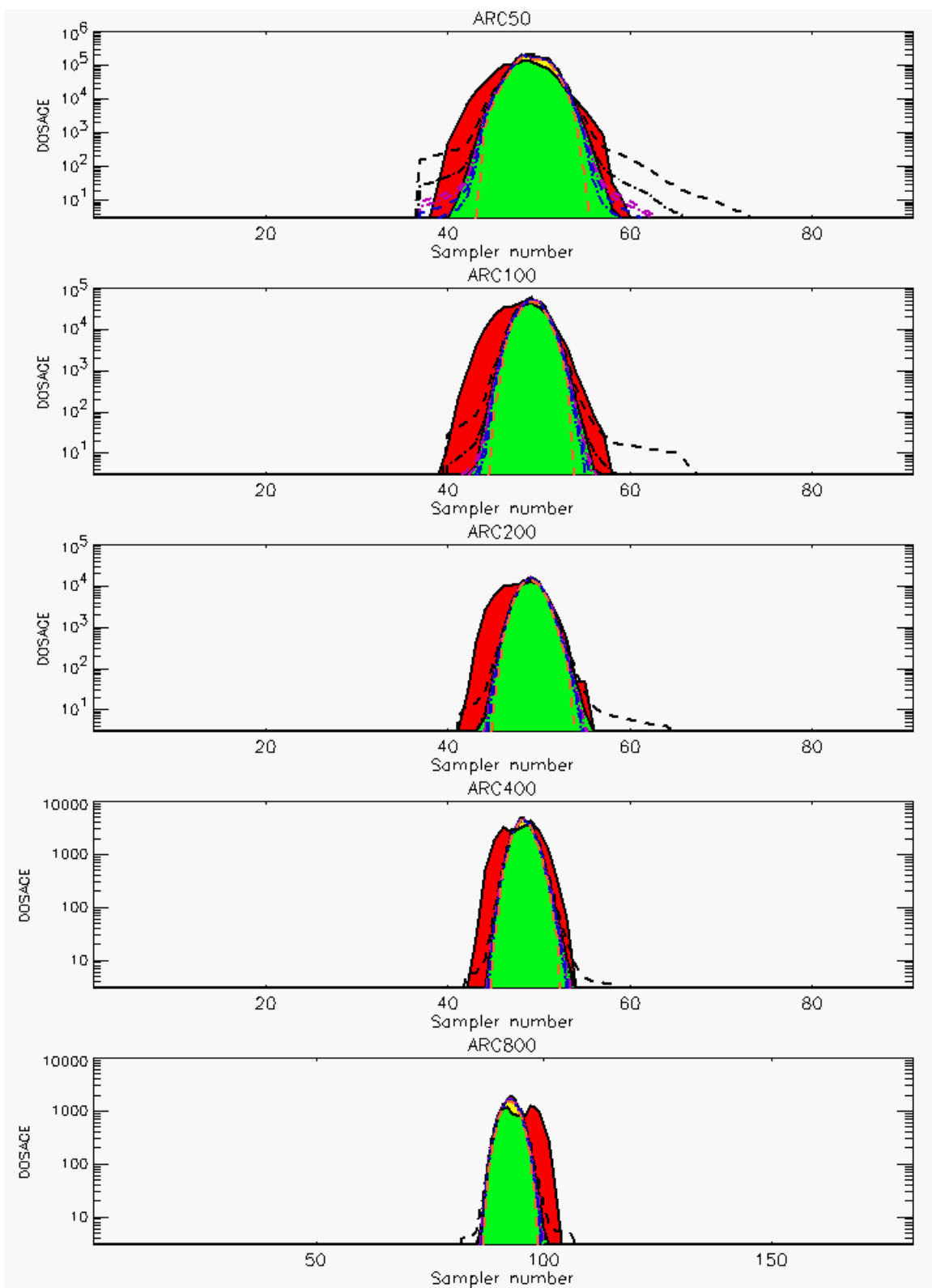


Figure C-29a. HPAC Probabilistic Prediction Outputs for Trial 37 on Linear Scale: Stability Class is 4



**Figure C-29b. HPAC Probabilistic Prediction Outputs for Trial 37 on Logarithmic Scale:
Stability Class is 4**

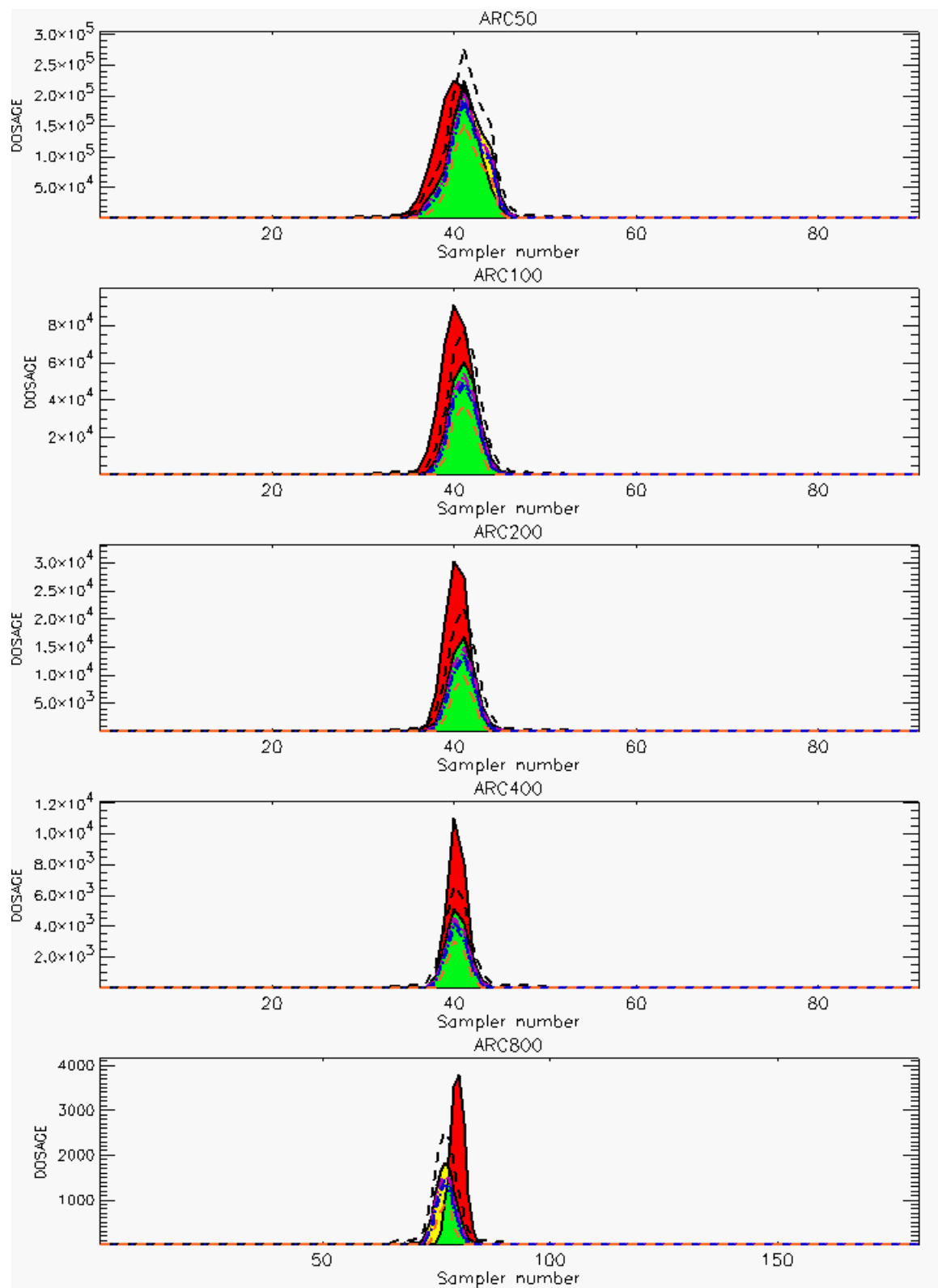


Figure C-30a. HPAC Probabilistic Prediction Outputs for Trial 38 on Linear Scale: Stability Class is 4

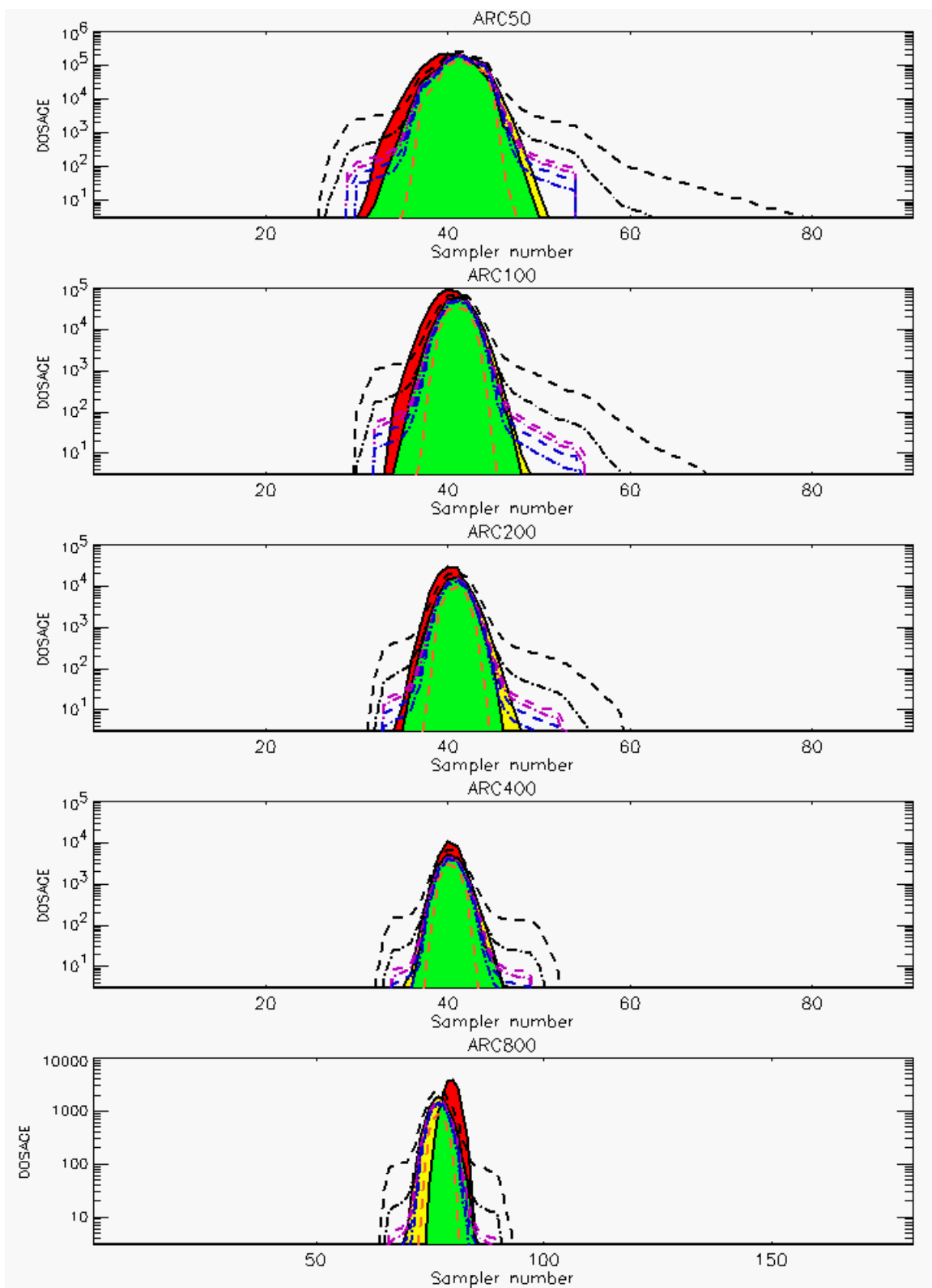


Figure C-30b. HPAC Probabilistic Prediction Outputs for Trial 38 on Logarithmic Scale: Stability Class is 4

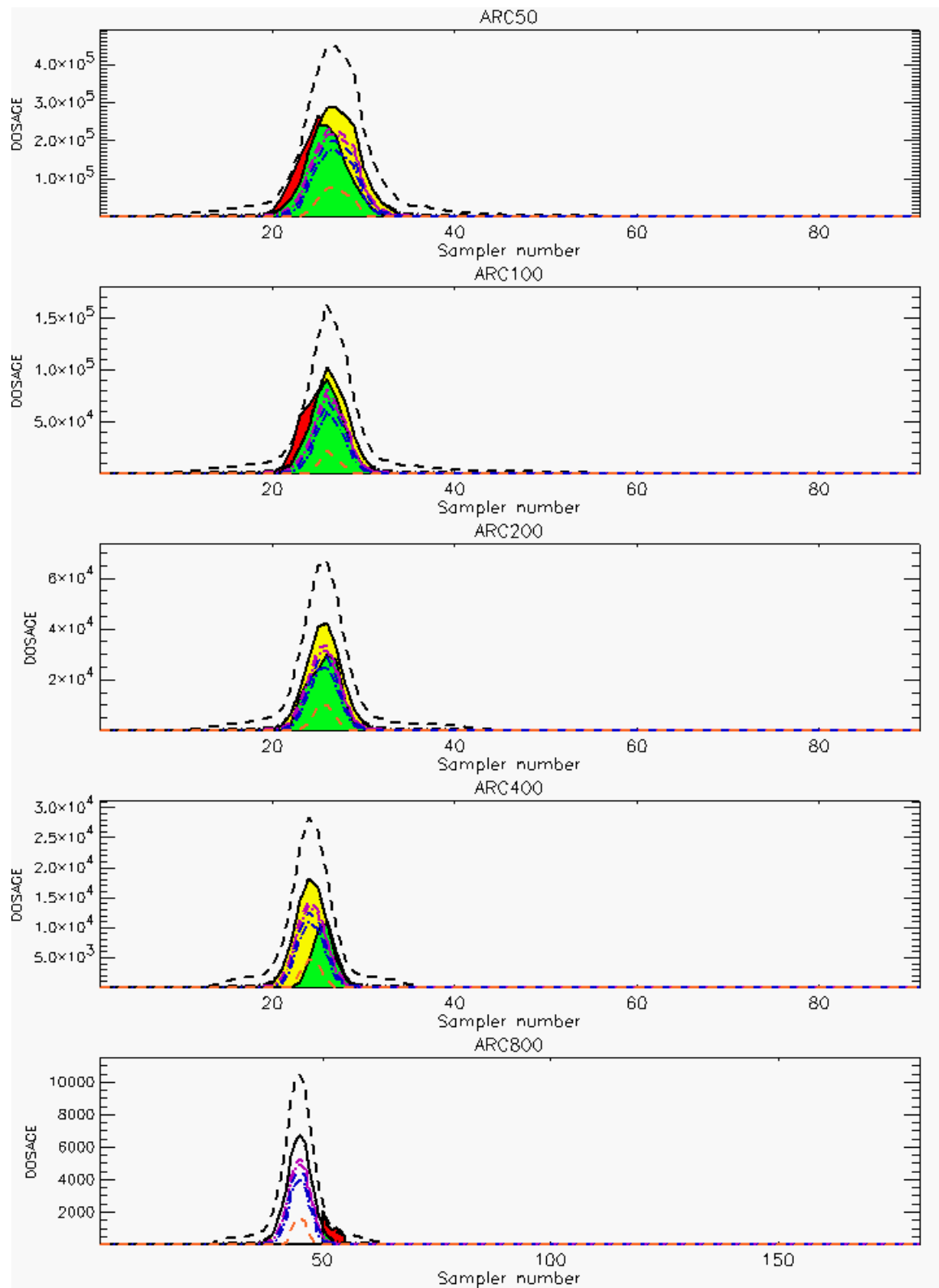


Figure C-31a. HPAC Probabilistic Prediction Outputs for Trial 39 on Linear Scale: Stability Class is 6 (Values for Samplers 40-49, 56-58 of 800-Meter Arc are Missing)

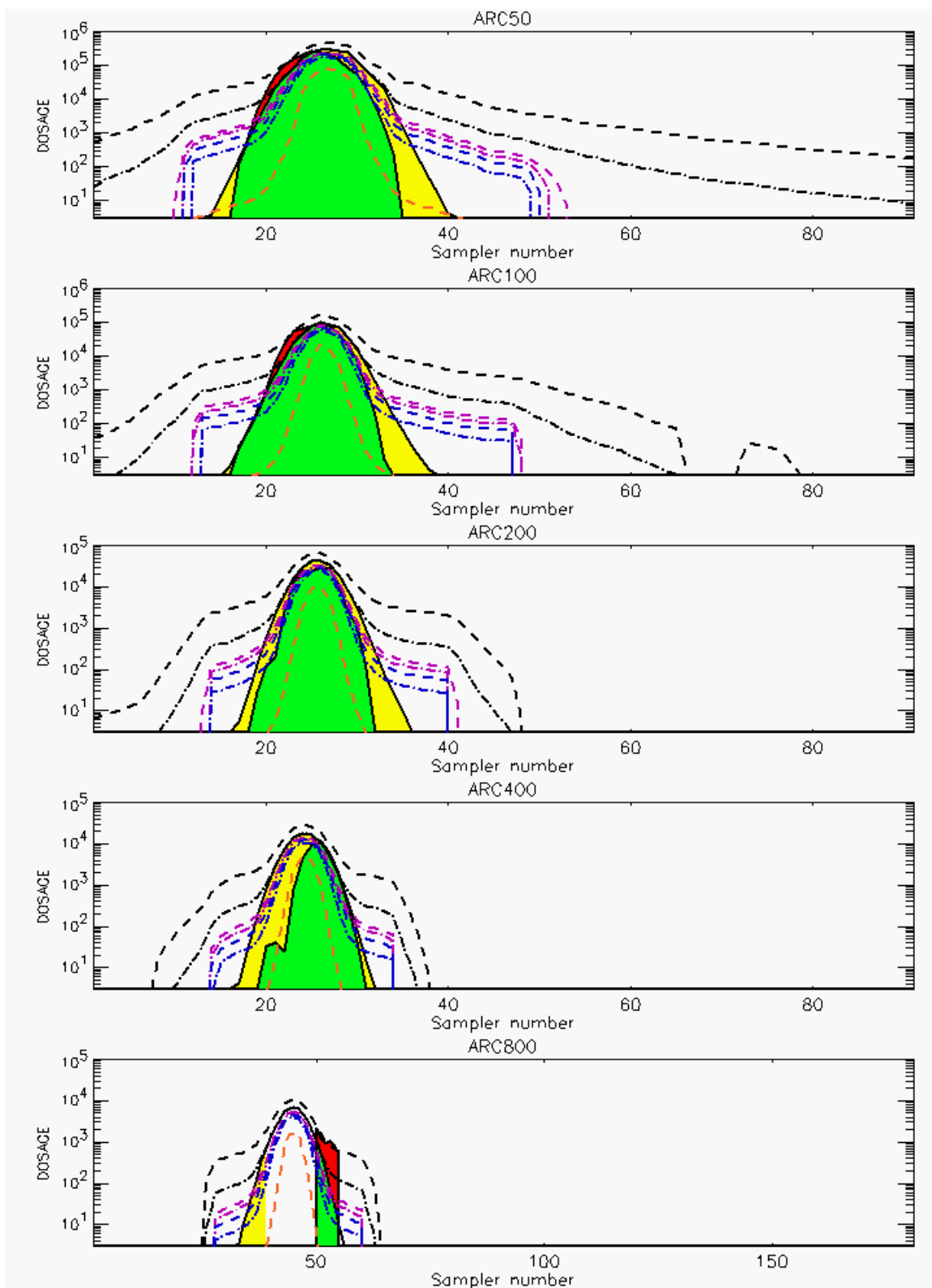


Figure C-31b. HPAC Probabilistic Prediction Outputs for Trial 39 on Logarithmic Scale: Stability Class is 6 (Values for Samplers 40-49, 56-58 of 800-Meter Arc are Missing)

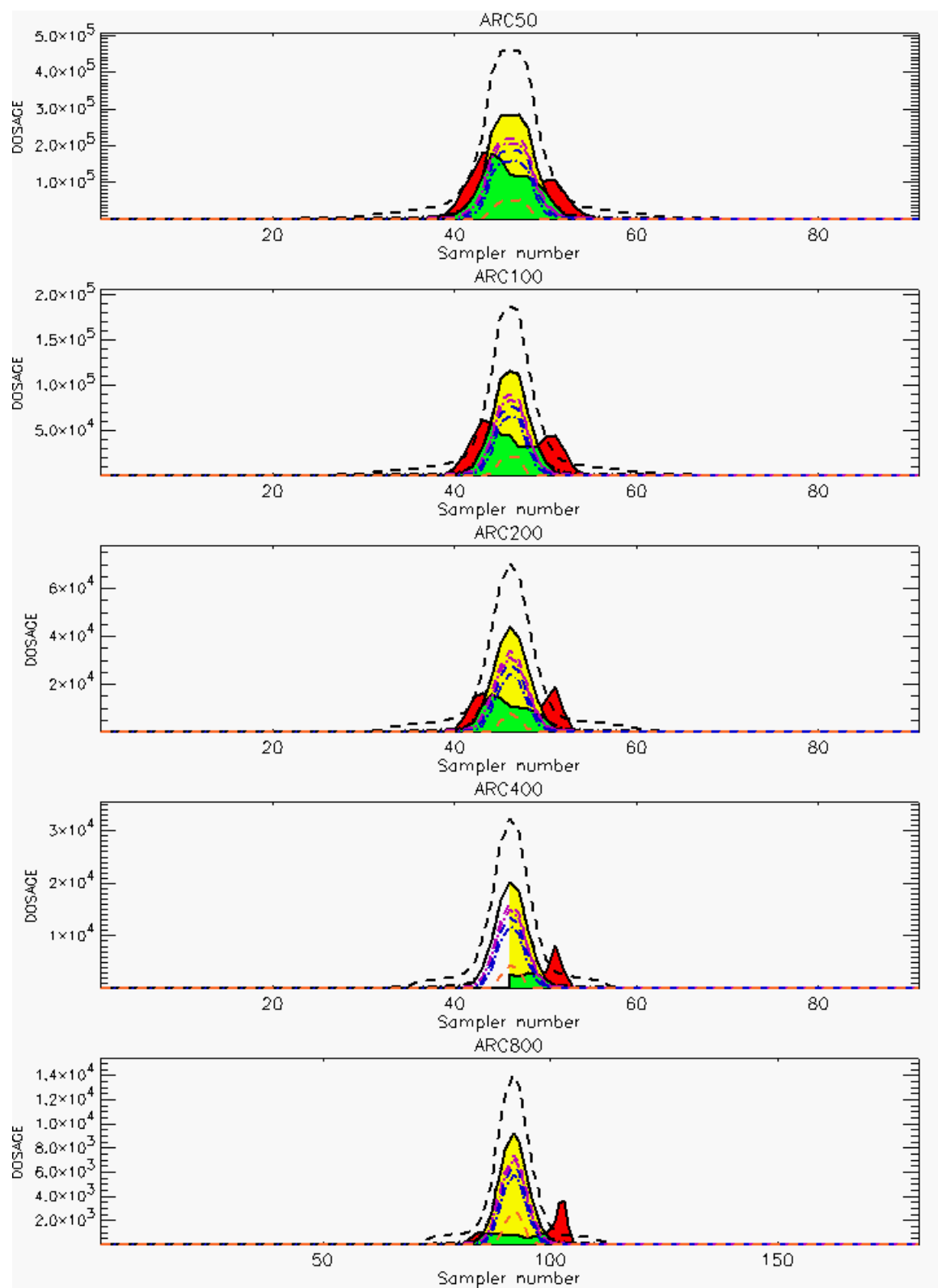


Figure C-32a. HPAC Probabilistic Prediction Outputs for Trial 40 on Linear Scale: Stability Class is 6 (Values for Samplers 39-45 of 400-Meter Arc are Missing)

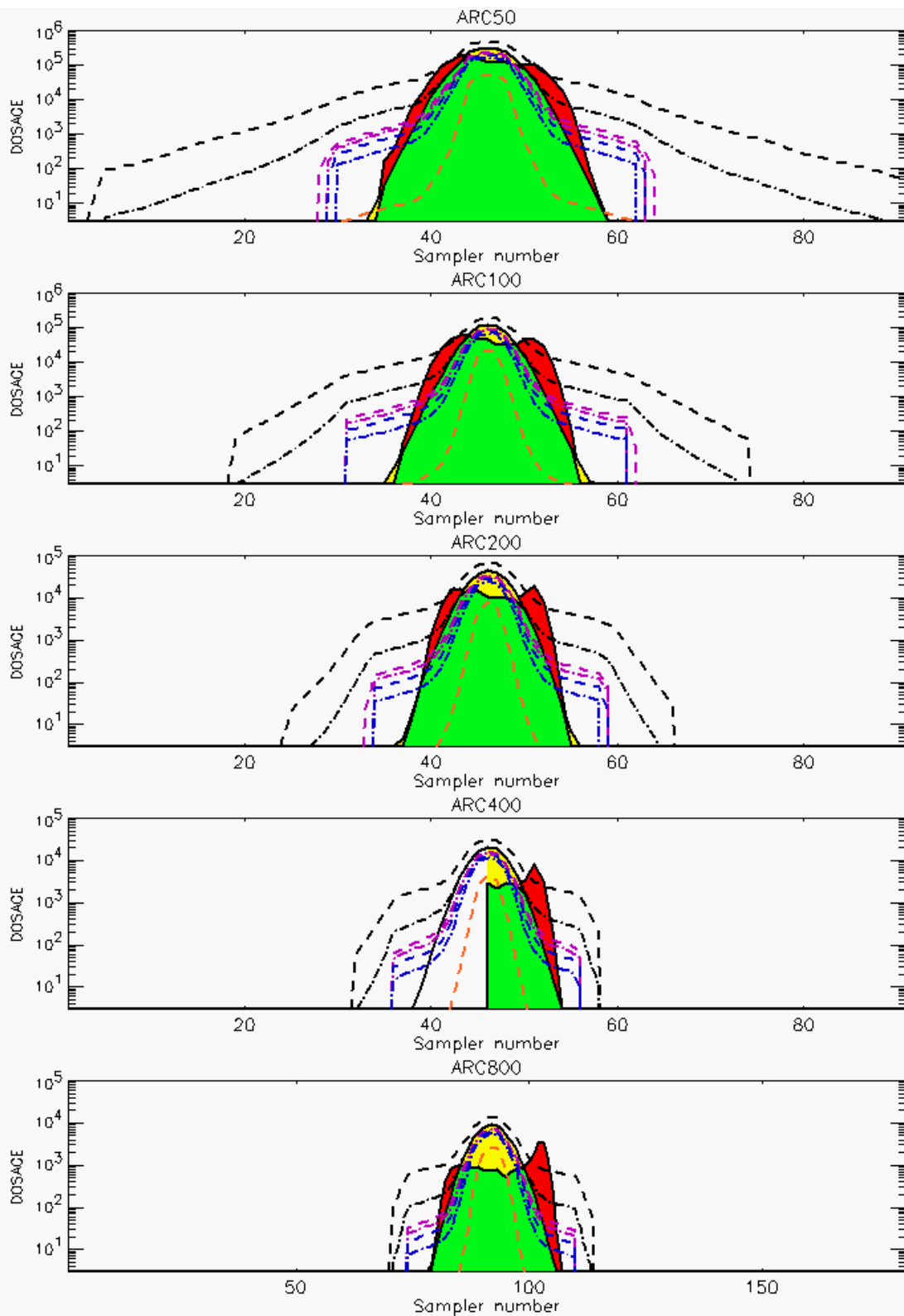


Figure C-32b. HPAC Probabilistic Prediction Outputs for Trial 40 on Logarithmic Scale: Stability Class is 6 (Values for Samplers 39-45 of 400-Meter Arc are Missing)

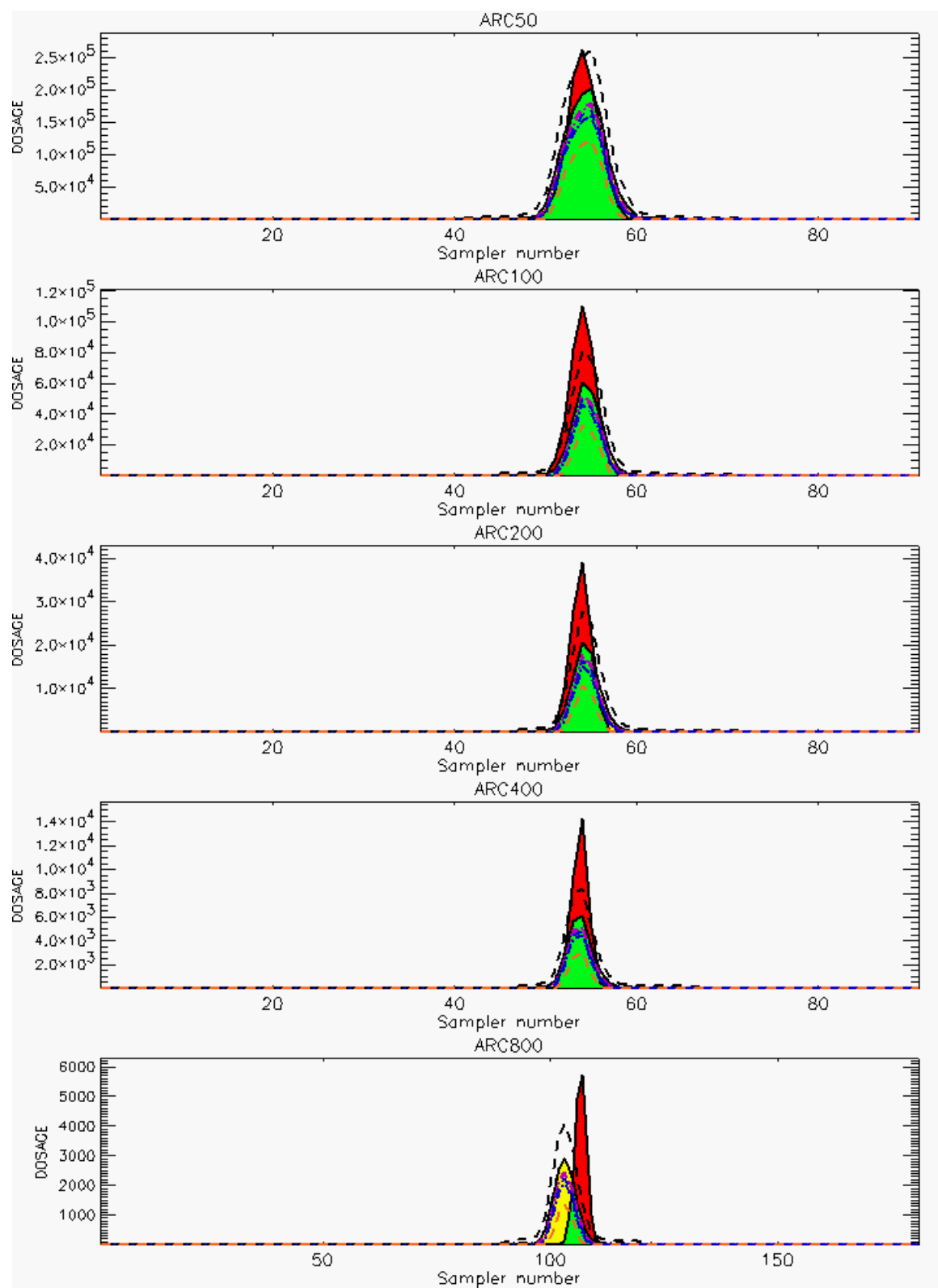
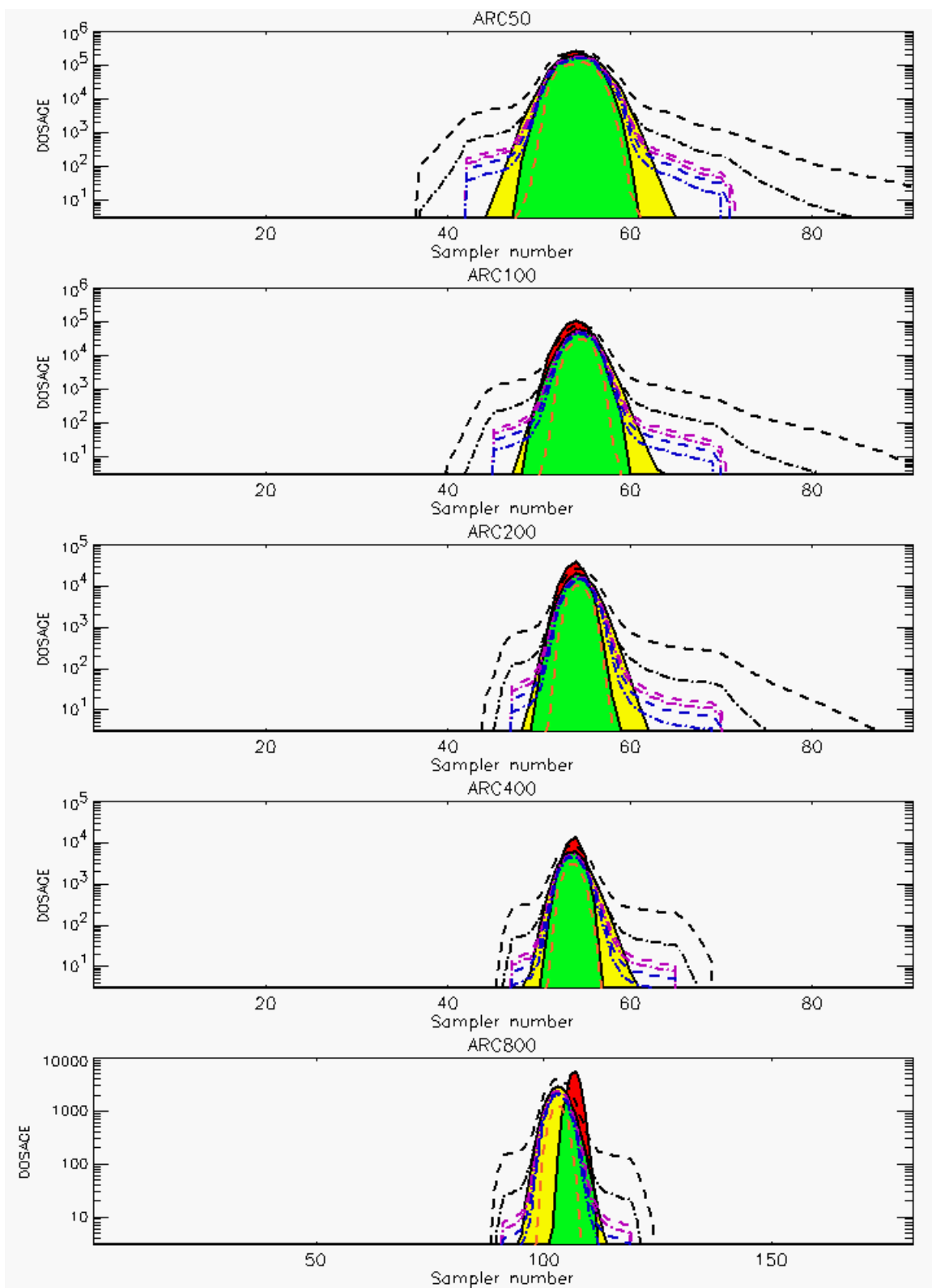


Figure C-33a. HPAC Probabilistic Prediction Outputs for Trial 41 on Linear Scale: Stability Class is 5



**Figure C-33b. HPAC Probabilistic Prediction Outputs for Trial 41 on Logarithmic Scale:
Stability Class is 5**

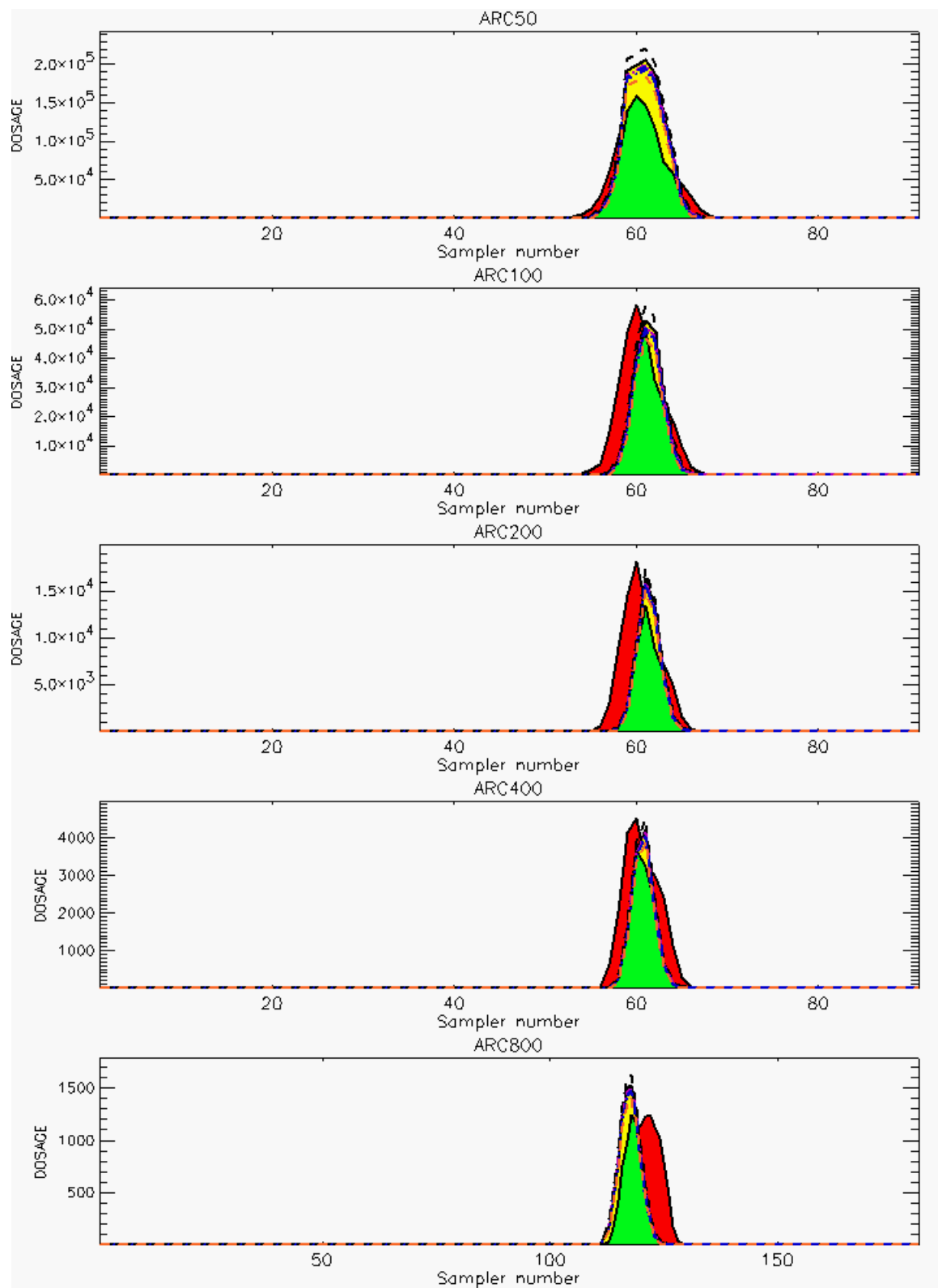


Figure C-34a. HPAC Probabilistic Prediction Outputs for Trial 42 on Linear Scale: Stability Class is 4

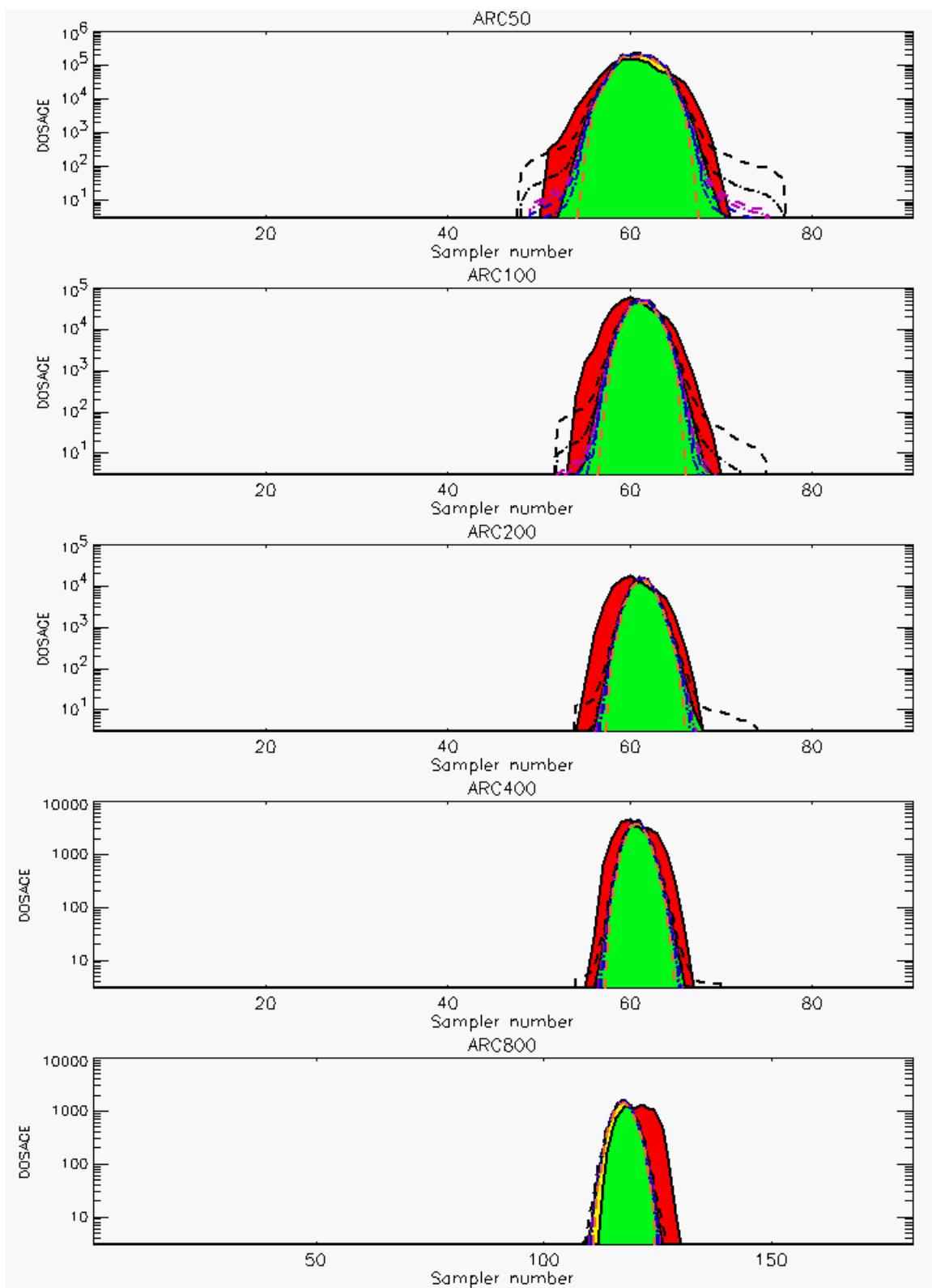


Figure C-34b. HPAC Probabilistic Prediction Outputs for Trial 42 on Logarithmic Scale: Stability Class is 4

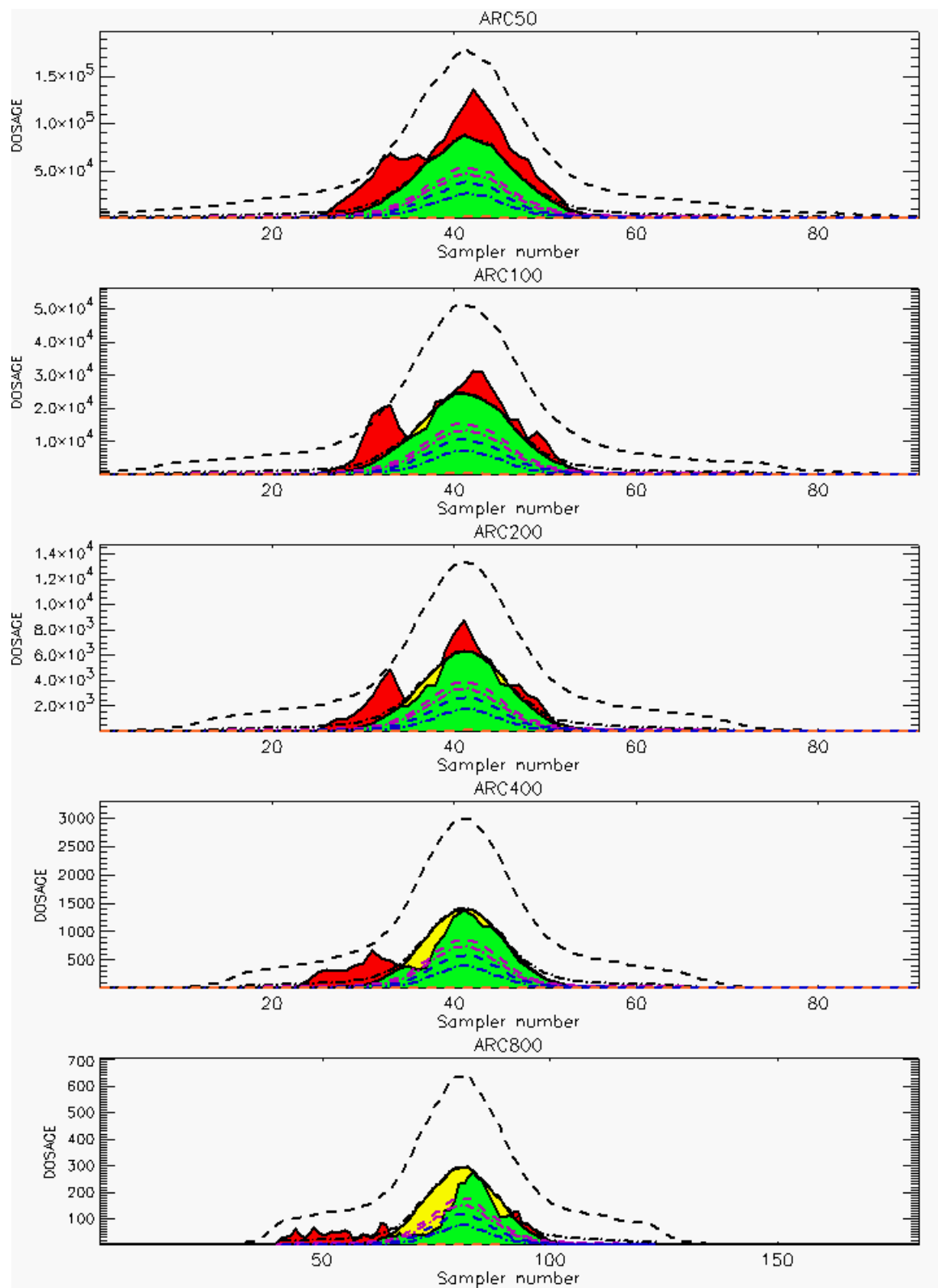


Figure C-35a. HPAC Probabilistic Prediction Outputs for Trial 43 on Linear Scale: Stability Class is 1

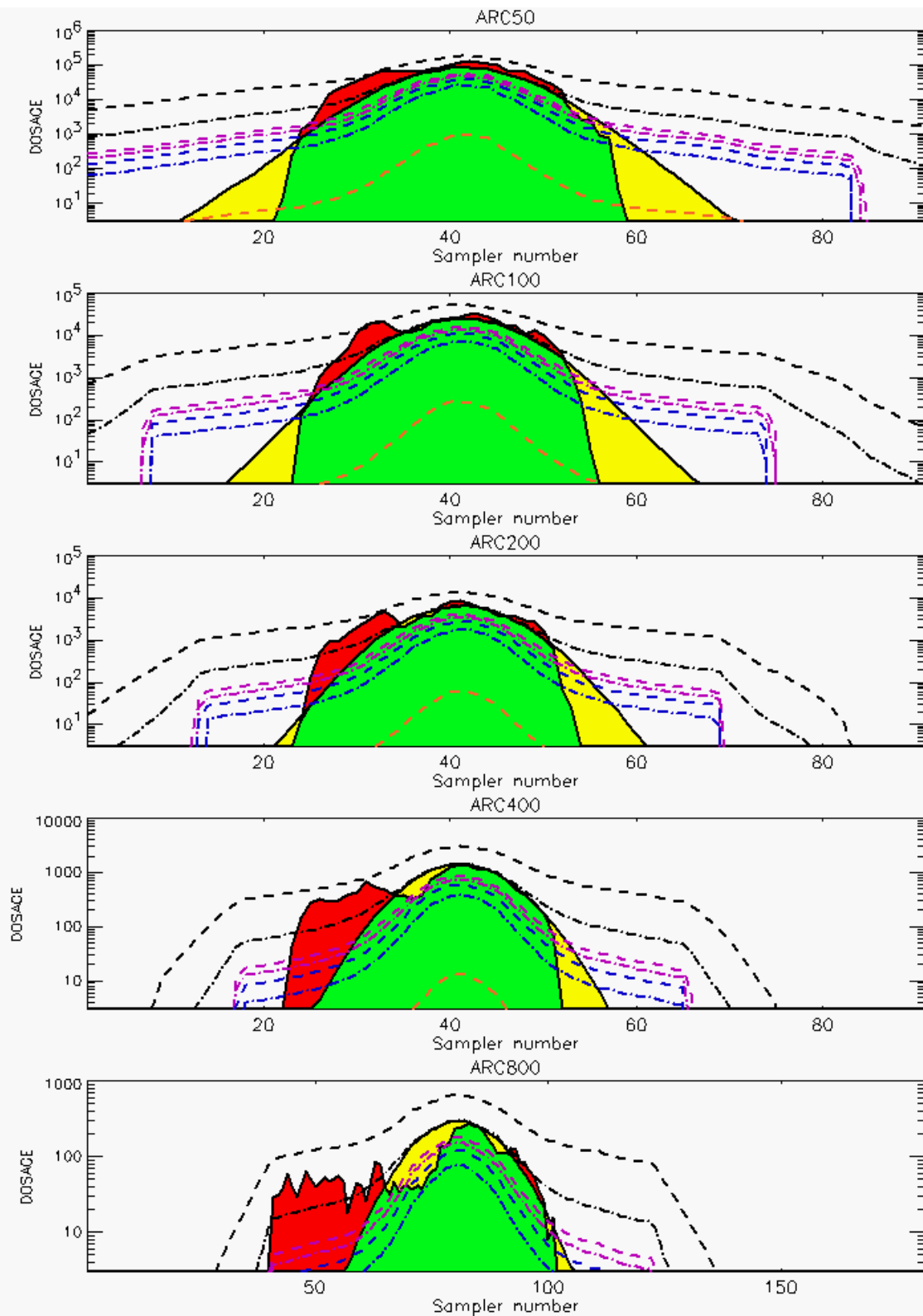


Figure C-35b. HPAC Probabilistic Prediction Outputs for Trial 43 on Logarithmic Scale: Stability Class is 1

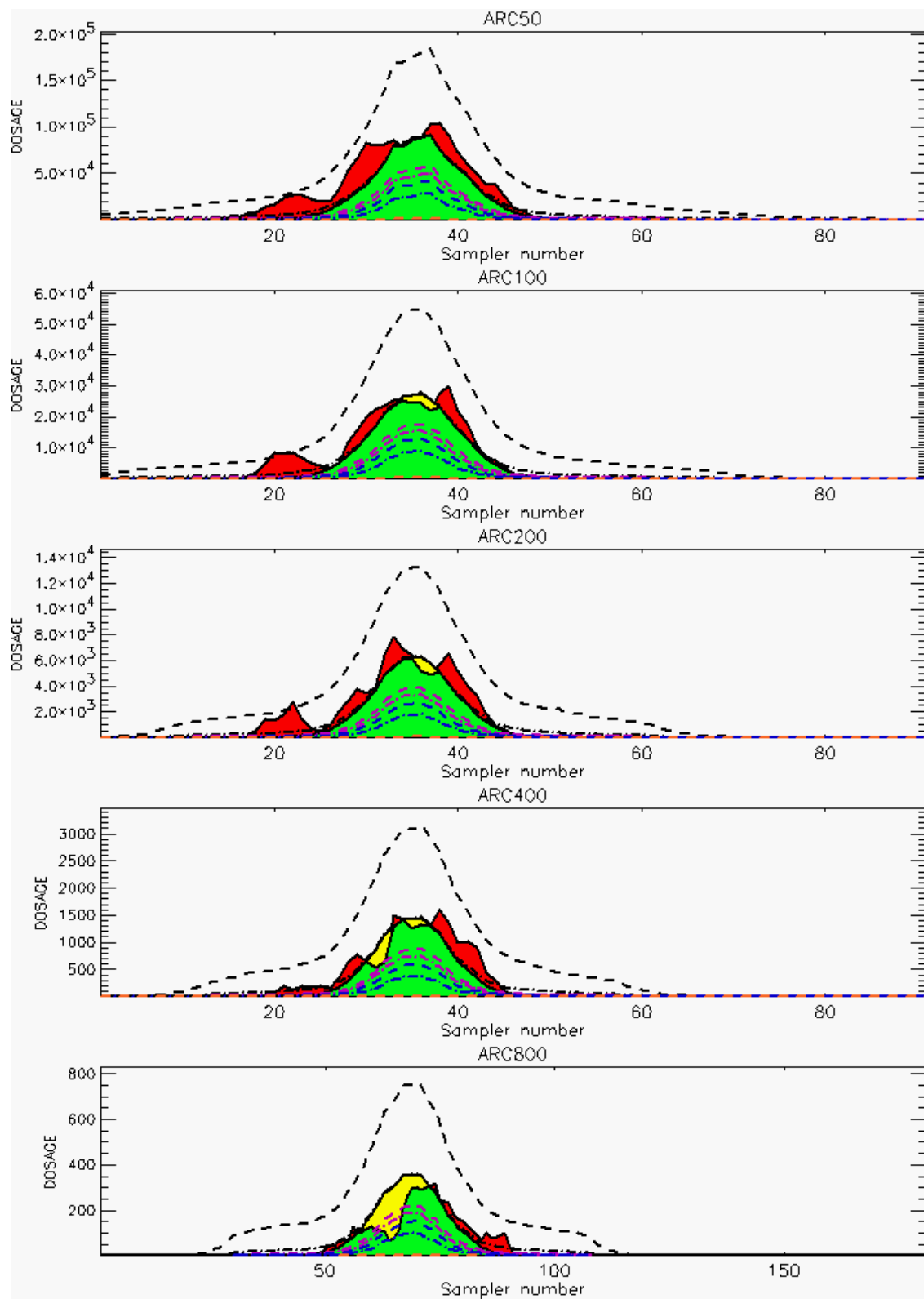


Figure C-36a. HPAC Probabilistic Prediction Outputs for Trial 44 on Linear Scale: Stability Class is 2

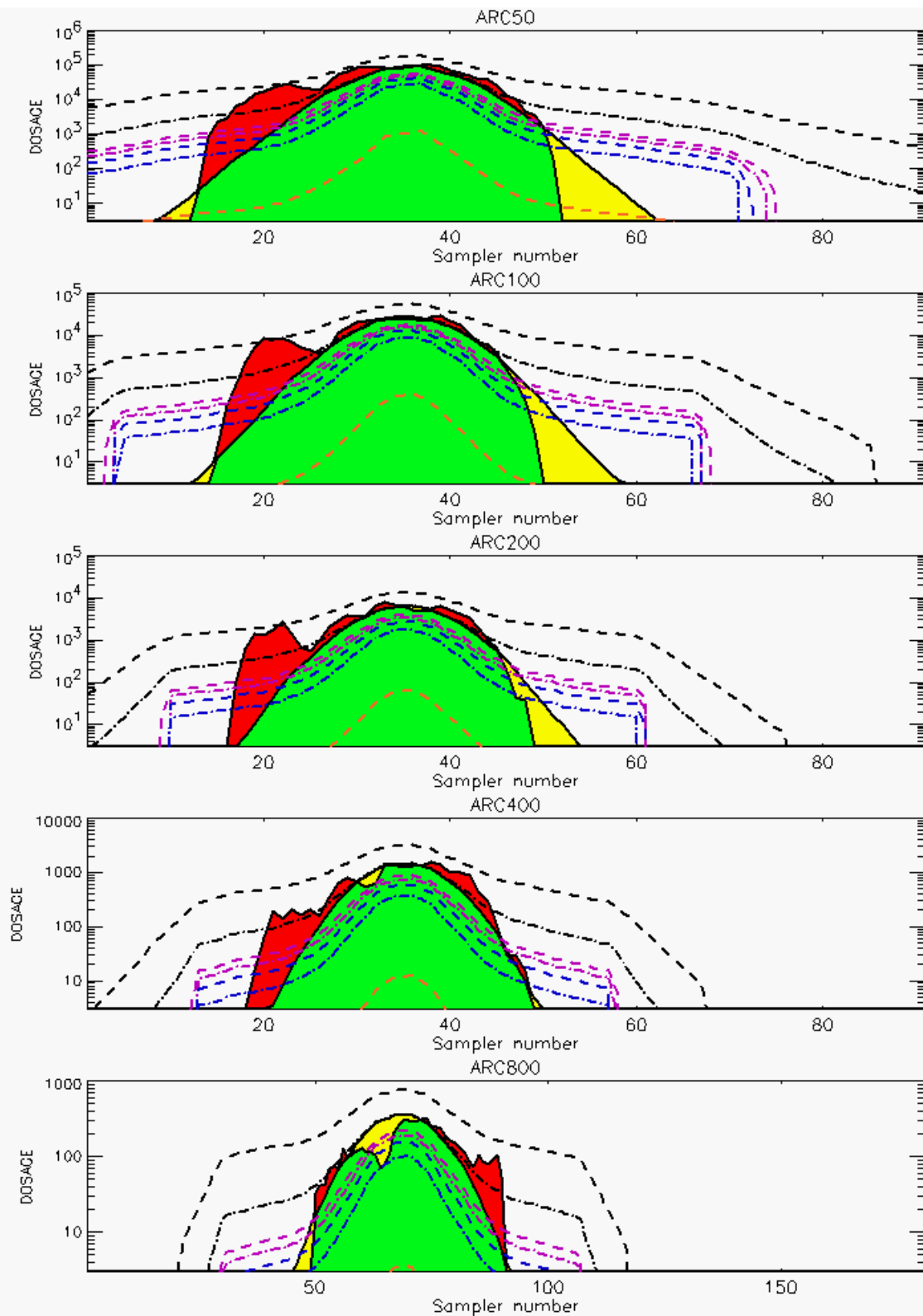


Figure C-36b. HPAC Probabilistic Prediction Outputs for Trial 44 on Logarithmic Scale: Stability Class is 2

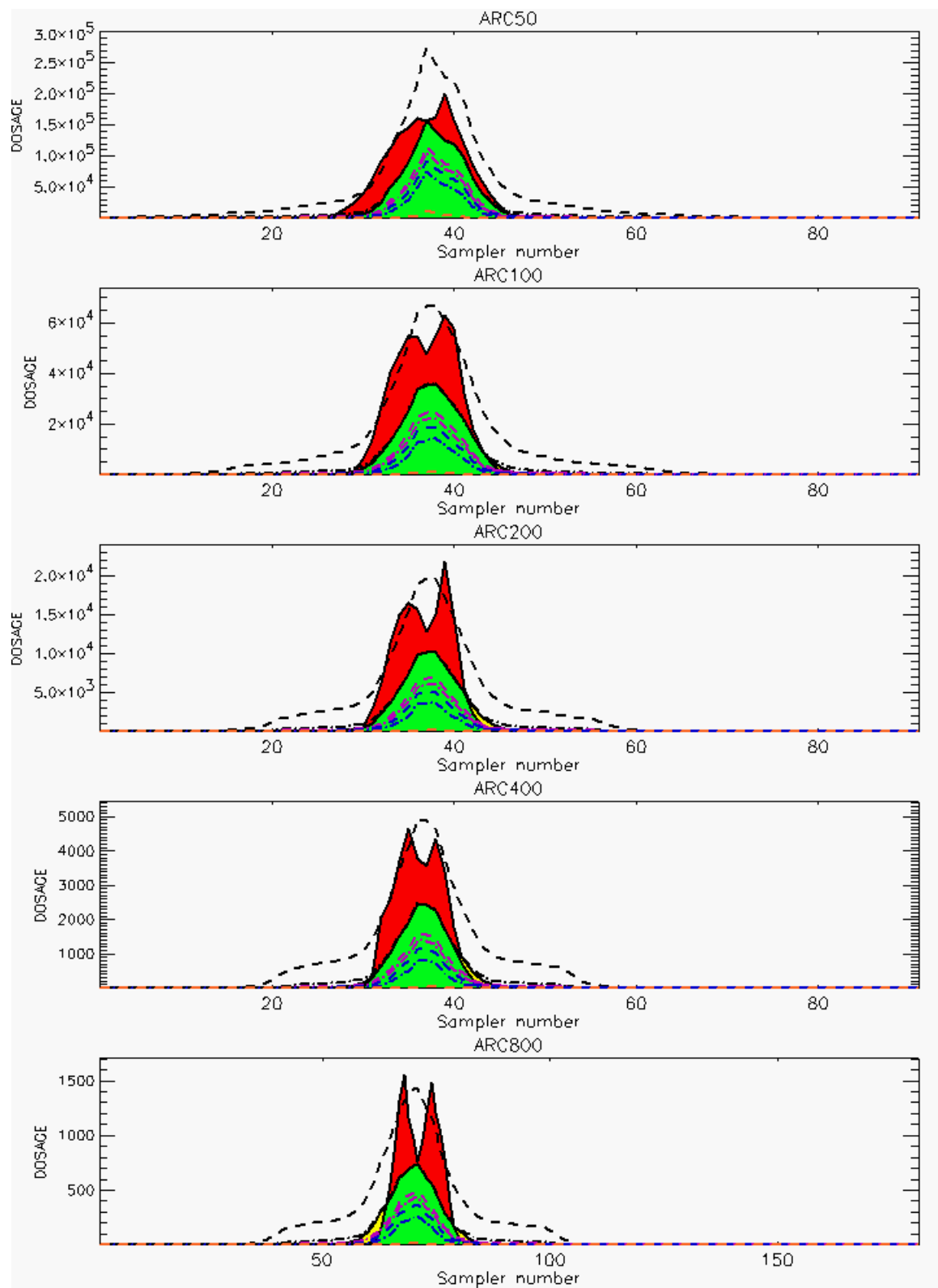


Figure C-37a. HPAC Probabilistic Prediction Outputs for Trial 45 on Linear Scale: Stability Class is 3

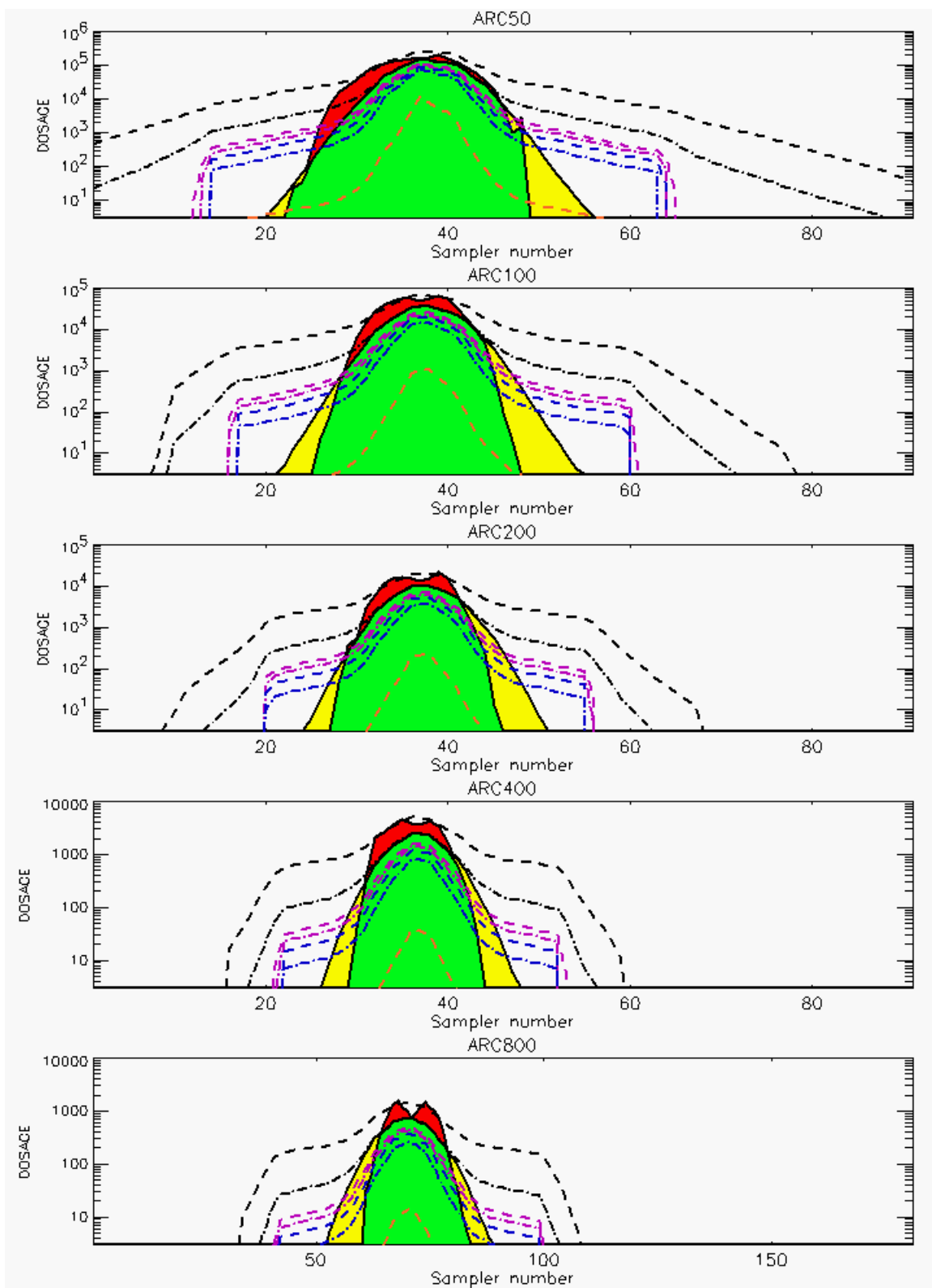


Figure C-37b. HPAC Probabilistic Prediction Outputs for Trial 45 on Logarithmic Scale: Stability Class is 3

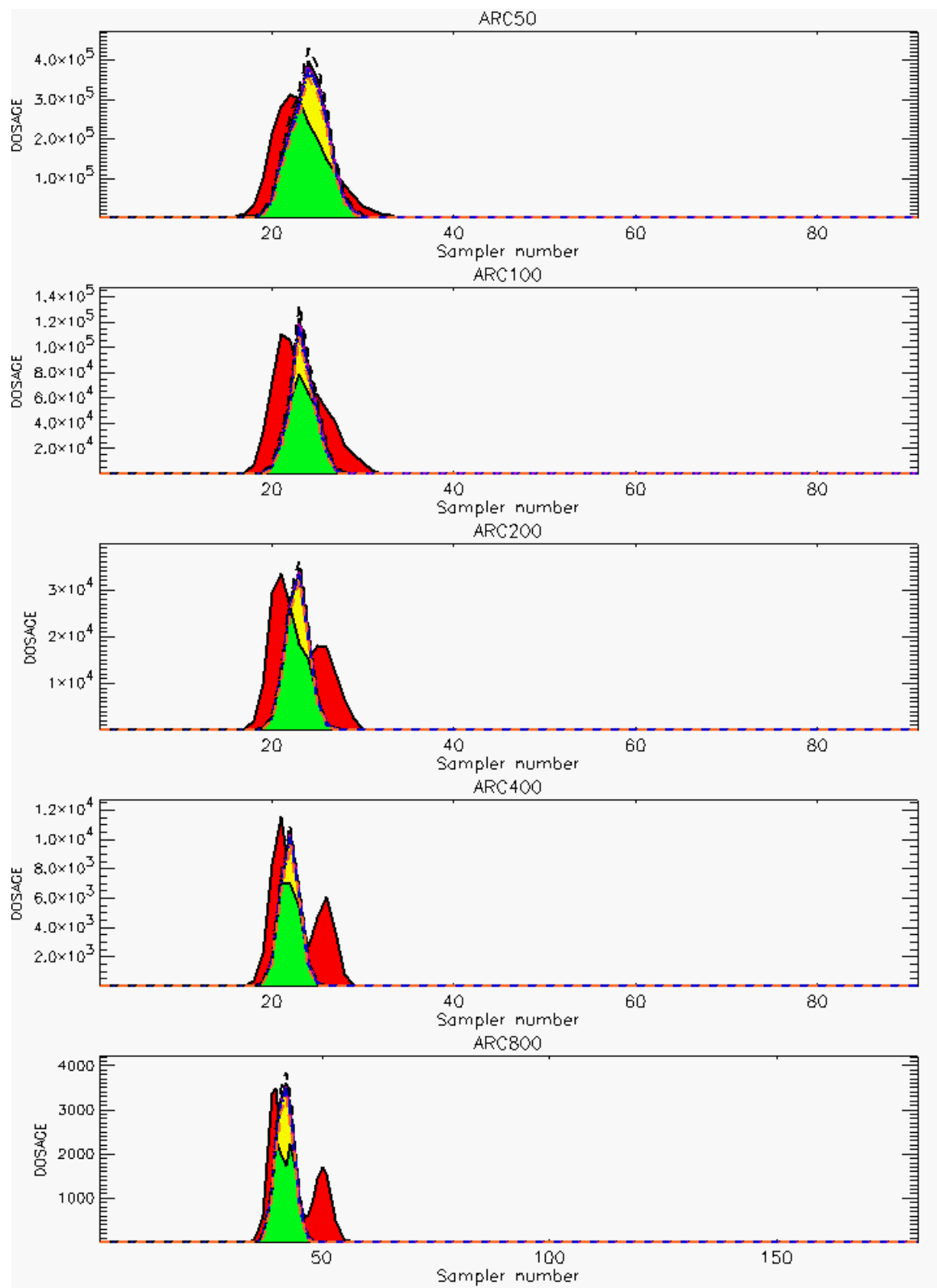
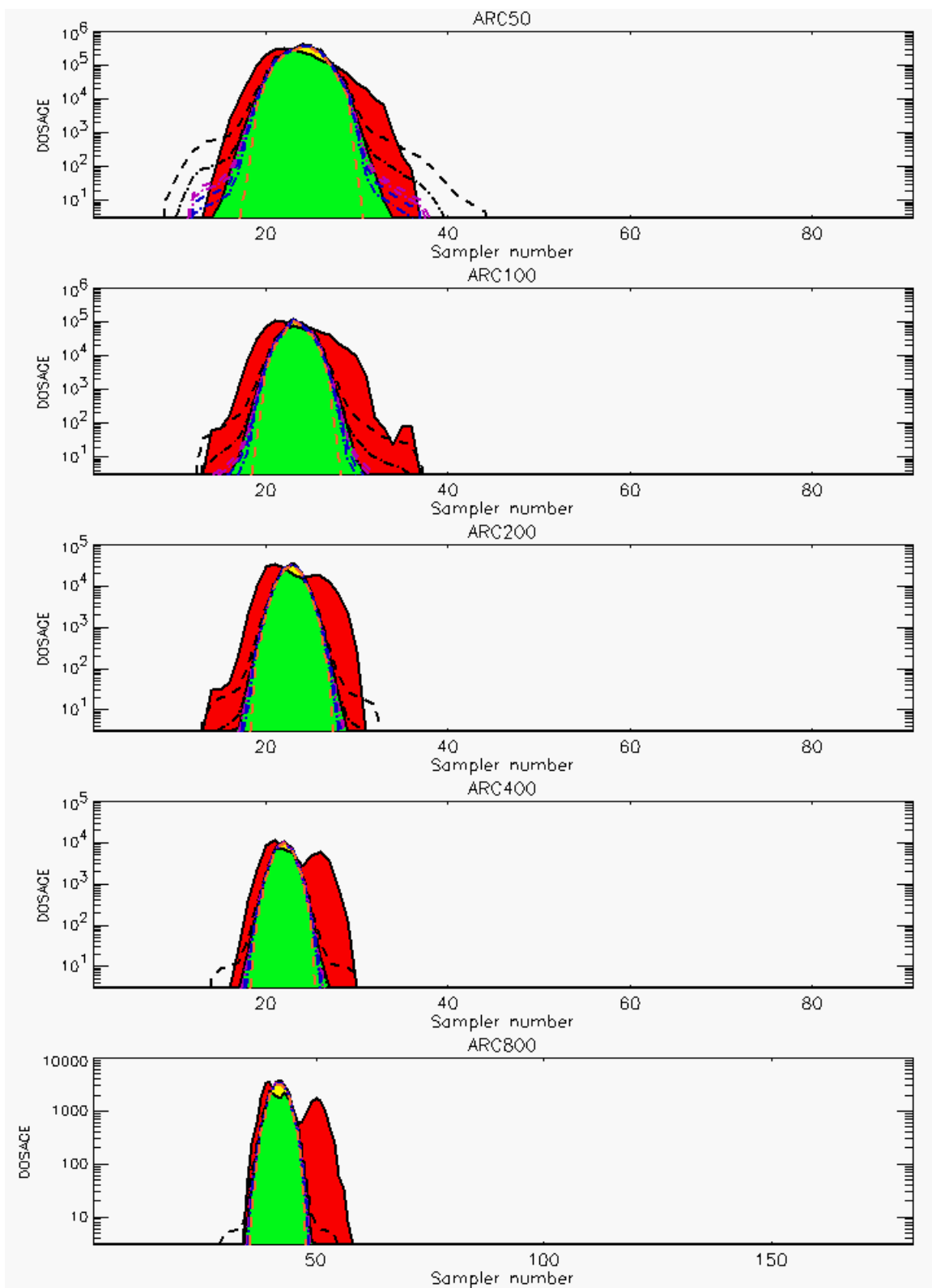


Figure C-38a. HPAC Probabilistic Prediction Outputs for Trial 46 on Linear Scale: Stability Class is 4



**Figure C-38b. HPAC Probabilistic Prediction Outputs for Trial 46 on Logarithmic Scale:
Stability Class is 4**

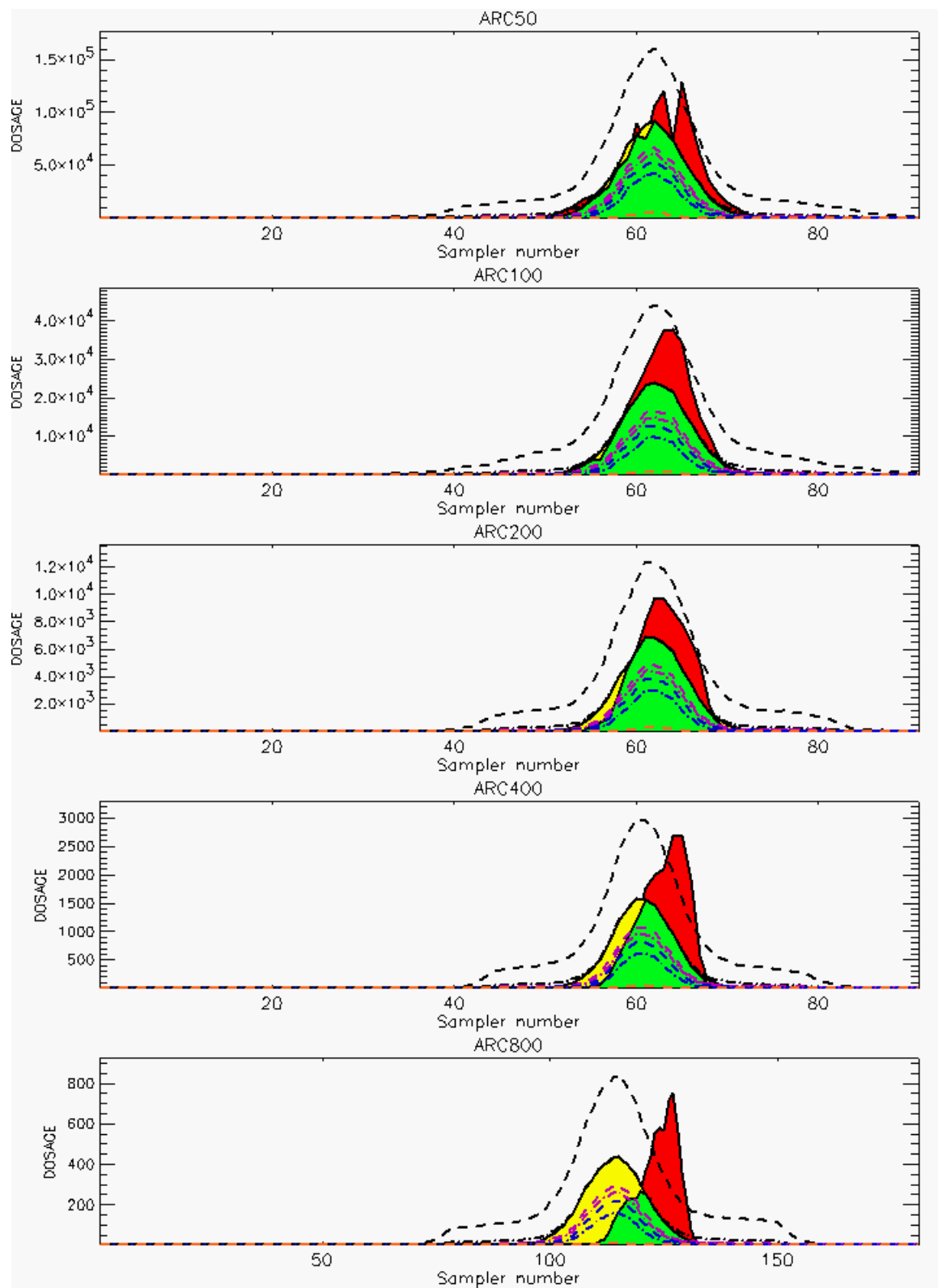


Figure C-39a. HPAC Probabilistic Prediction Outputs for Trial 48 on Linear Scale: Stability Class is 3

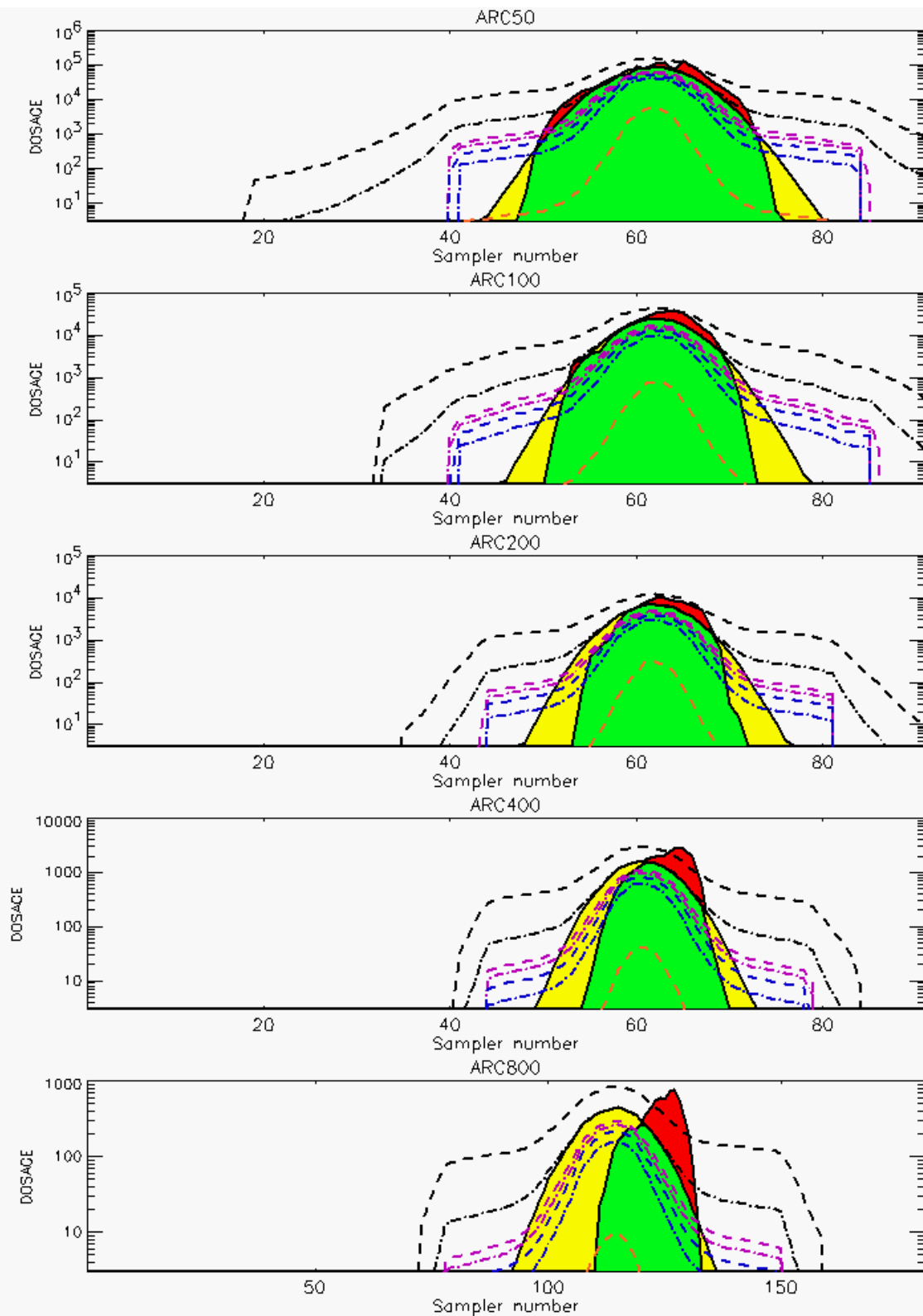


Figure C-39b. HPAC Probabilistic Prediction Outputs for Trial 48 on Logarithmic Scale: Stability Class is 3

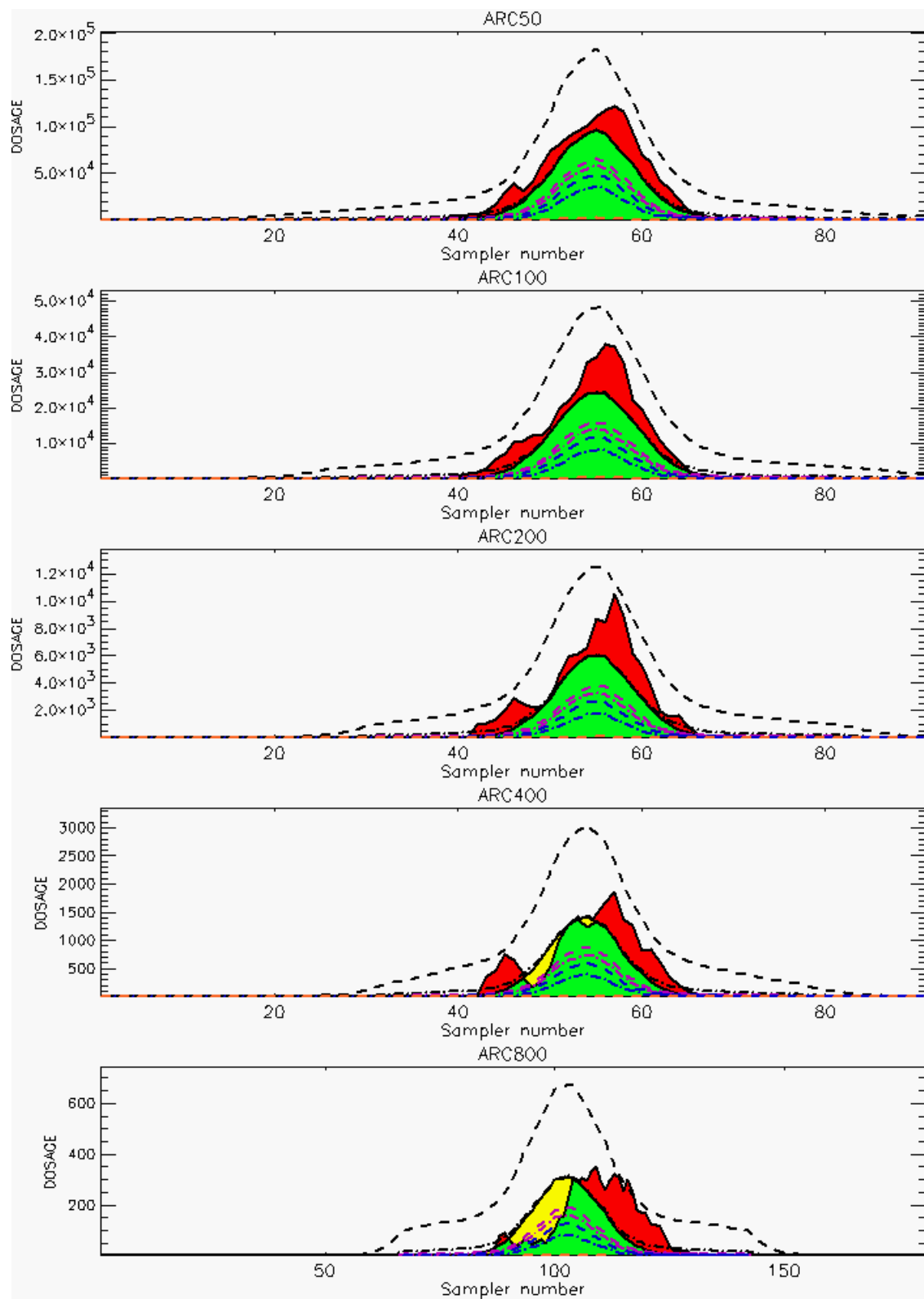


Figure C-40a. HPAC Probabilistic Prediction Outputs for Trial 49 on Linear Scale: Stability Class is 2

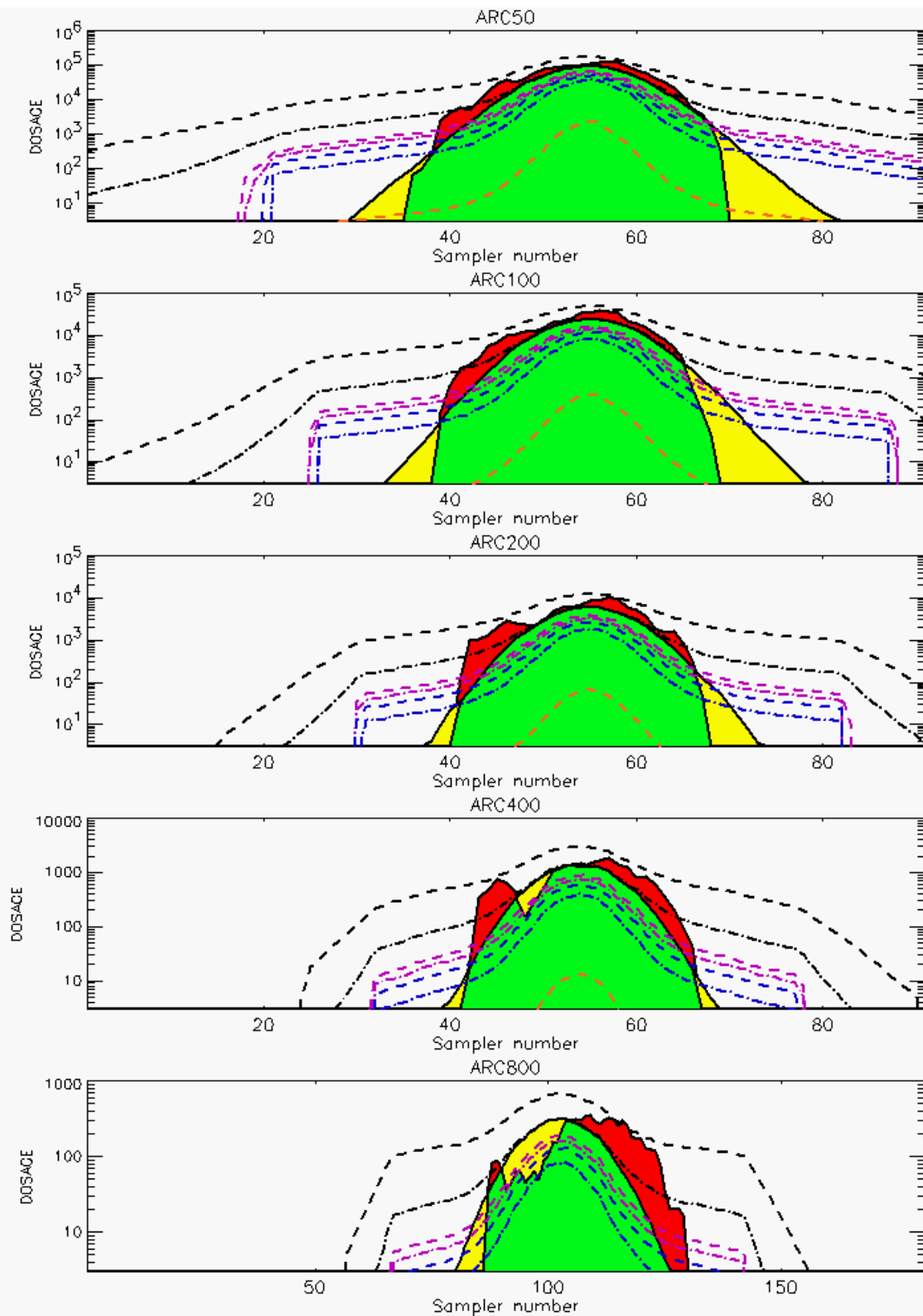


Figure C-40b. HPAC Probabilistic Prediction Outputs for Trial 49 on Logarithmic Scale: Stability Class is 2

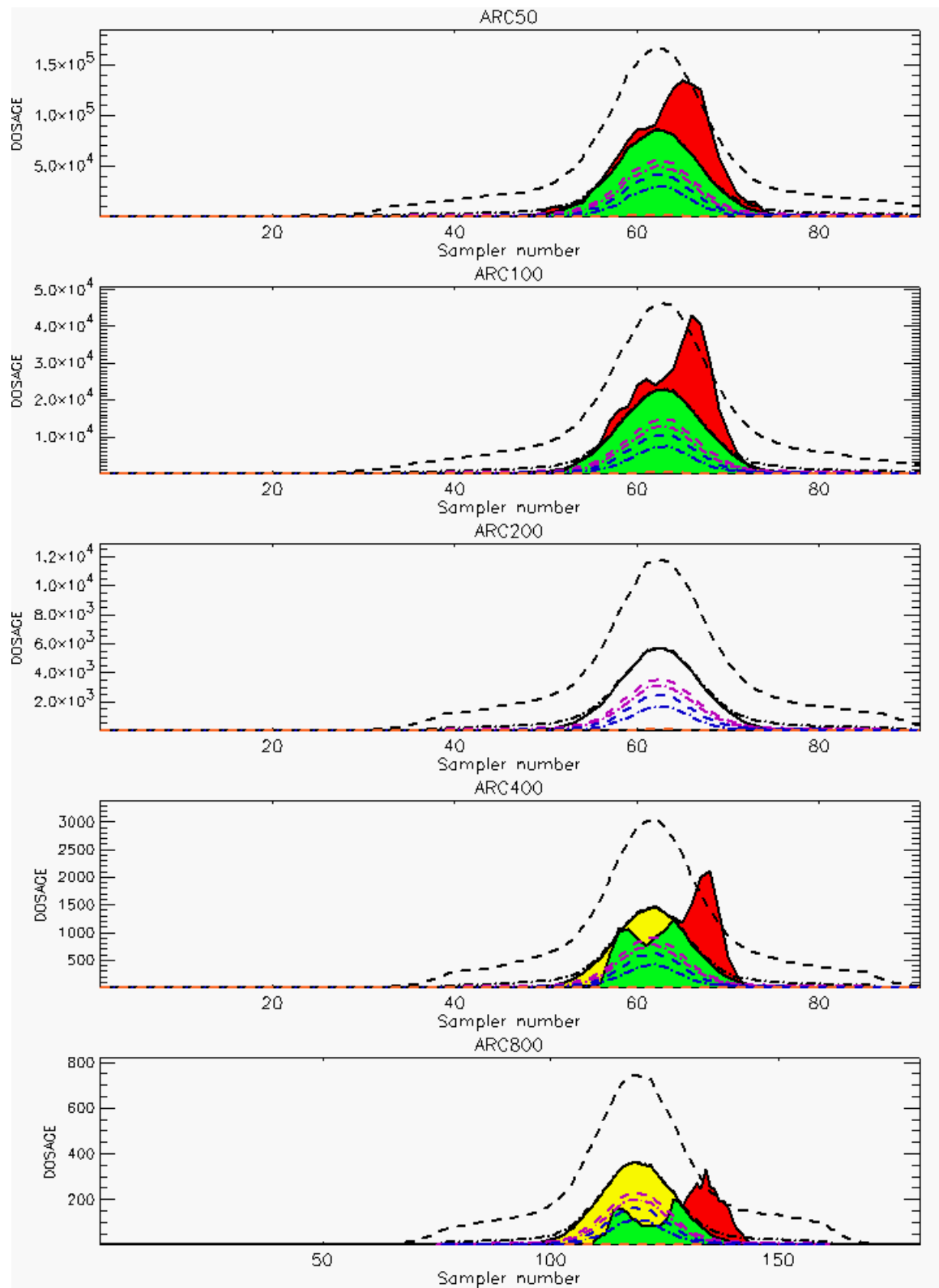


Figure C-41a. HPAC Probabilistic Prediction Outputs for Trial 50 on Linear Scale: Stability Class is 2 (Values for Samplers of 200-Meter Arc are Replaced by Missing Values [Ref. C-2])

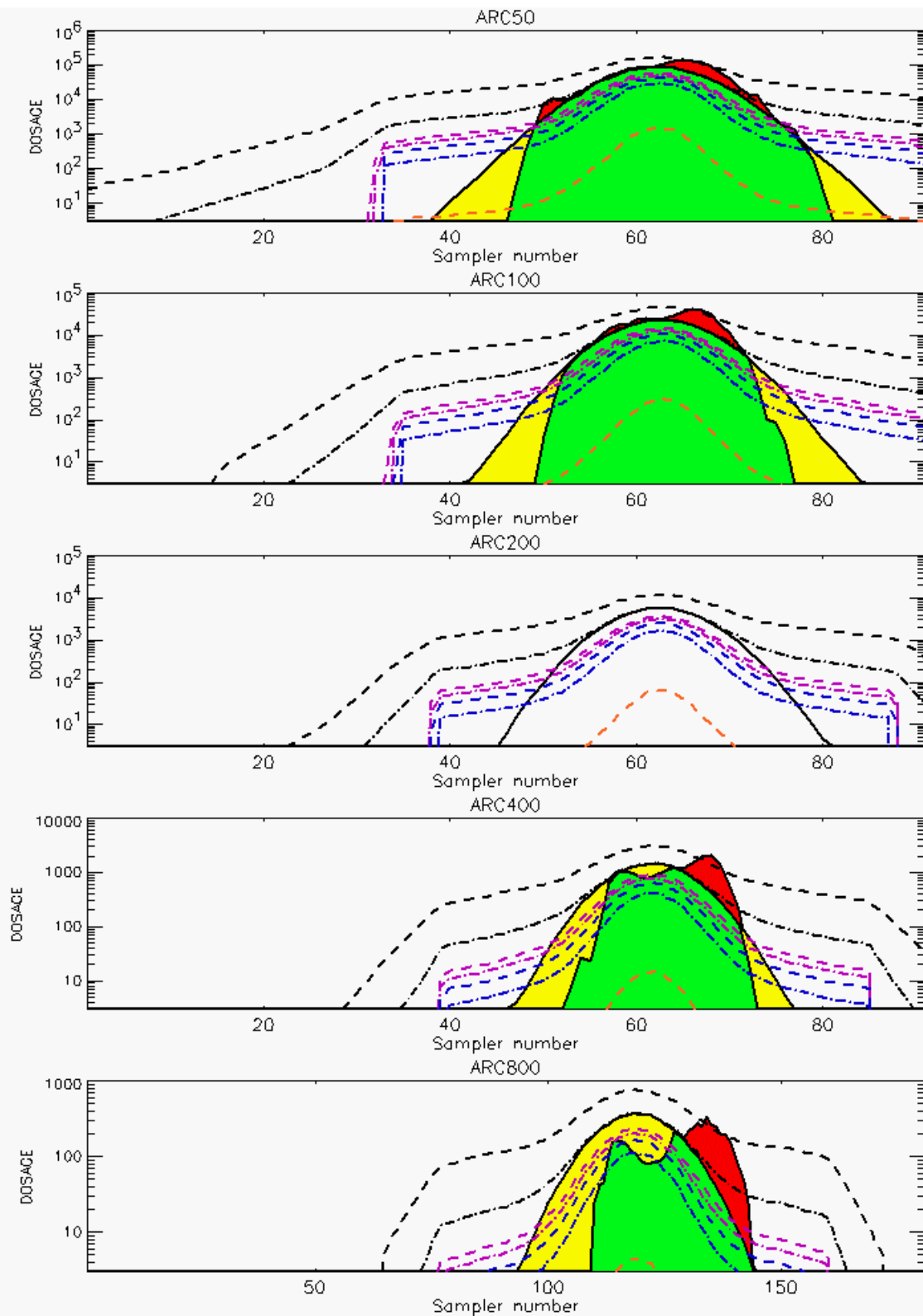


Figure C-41b. HPAC Probabilistic Prediction Outputs for Trial 50 on Logarithmic Scale: Stability Class is 2 (Values for Samplers of 200-Meter Arc are Replaced by Missing Values [Ref. C-2])

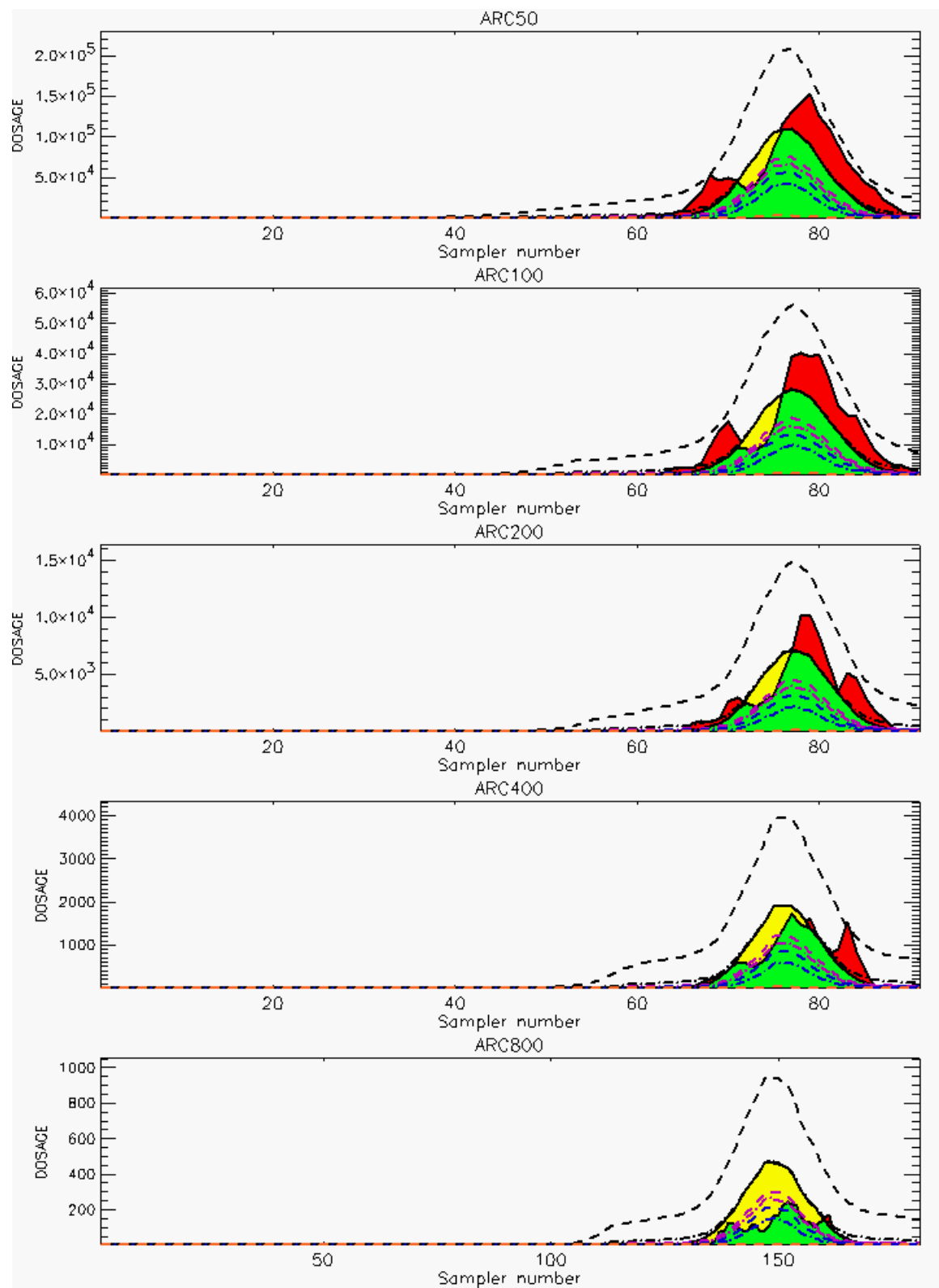


Figure C-42a. HPAC Probabilistic Prediction Outputs for Trial 51 on Linear Scale: Stability Class is 2

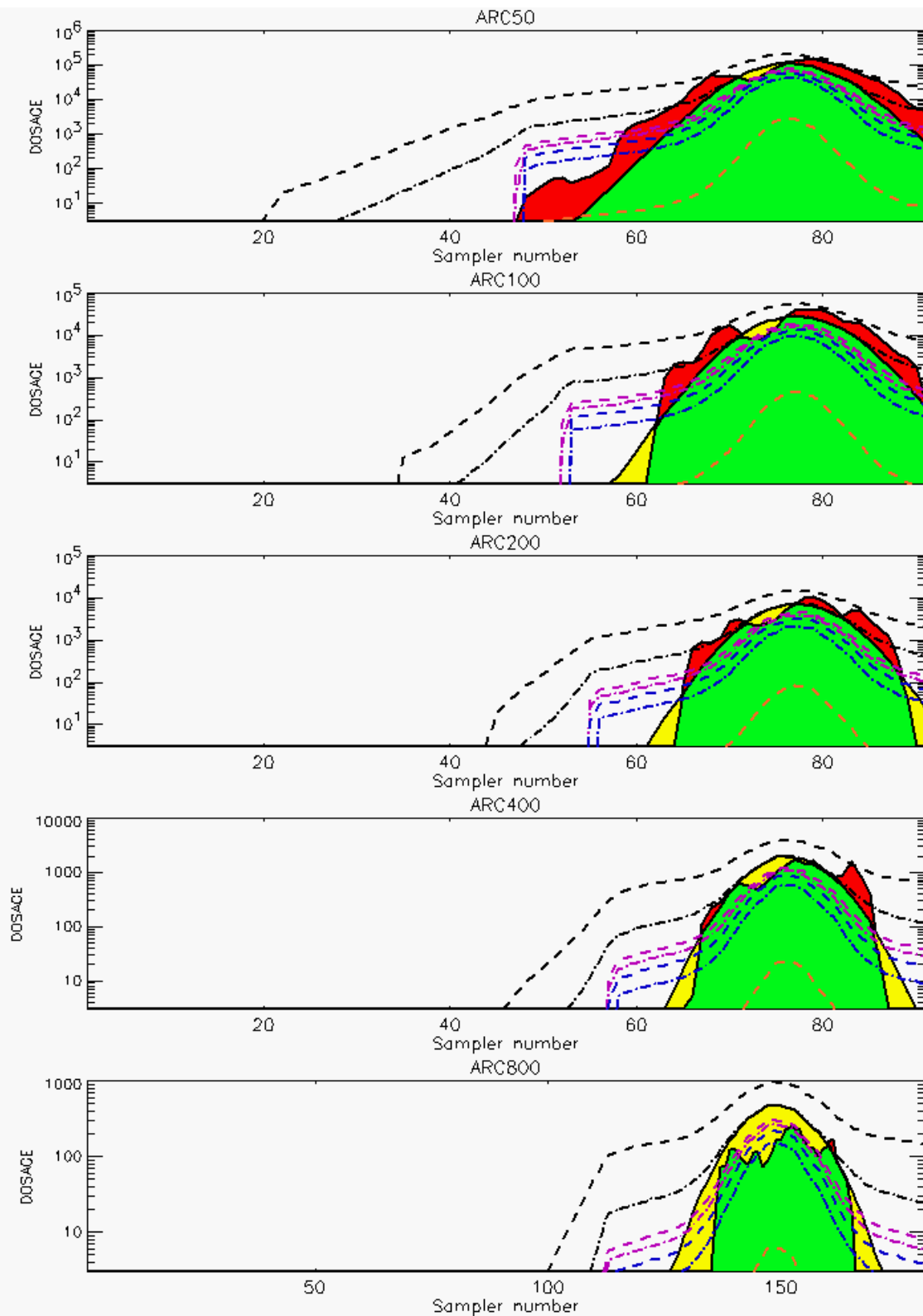


Figure C-42b. HPAC Probabilistic Prediction Outputs for Trial 51 on Logarithmic Scale: Stability Class is 2

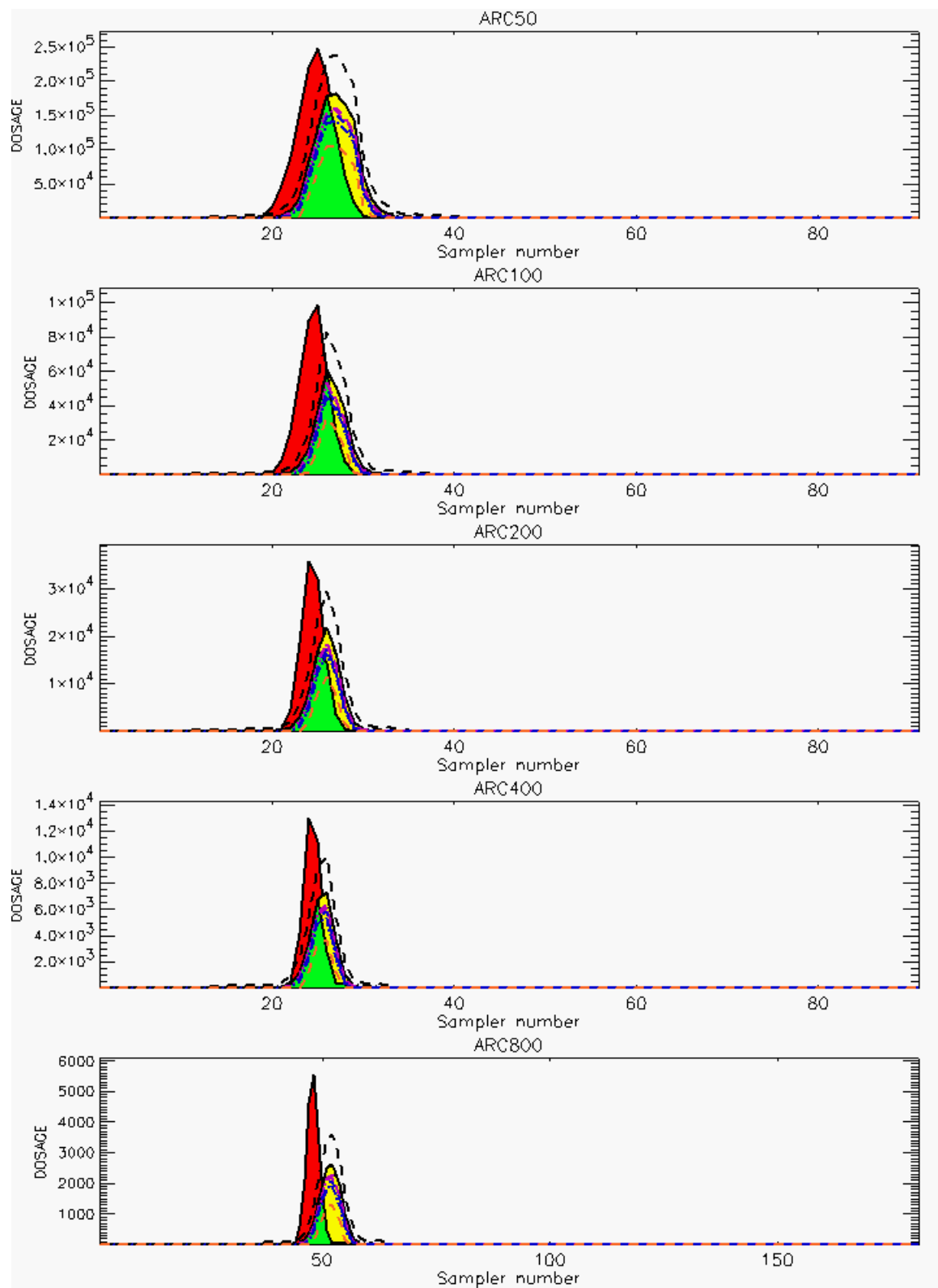


Figure C-43a. HPAC Probabilistic Prediction Outputs for Trial 54 on Linear Scale: Stability Class is 5

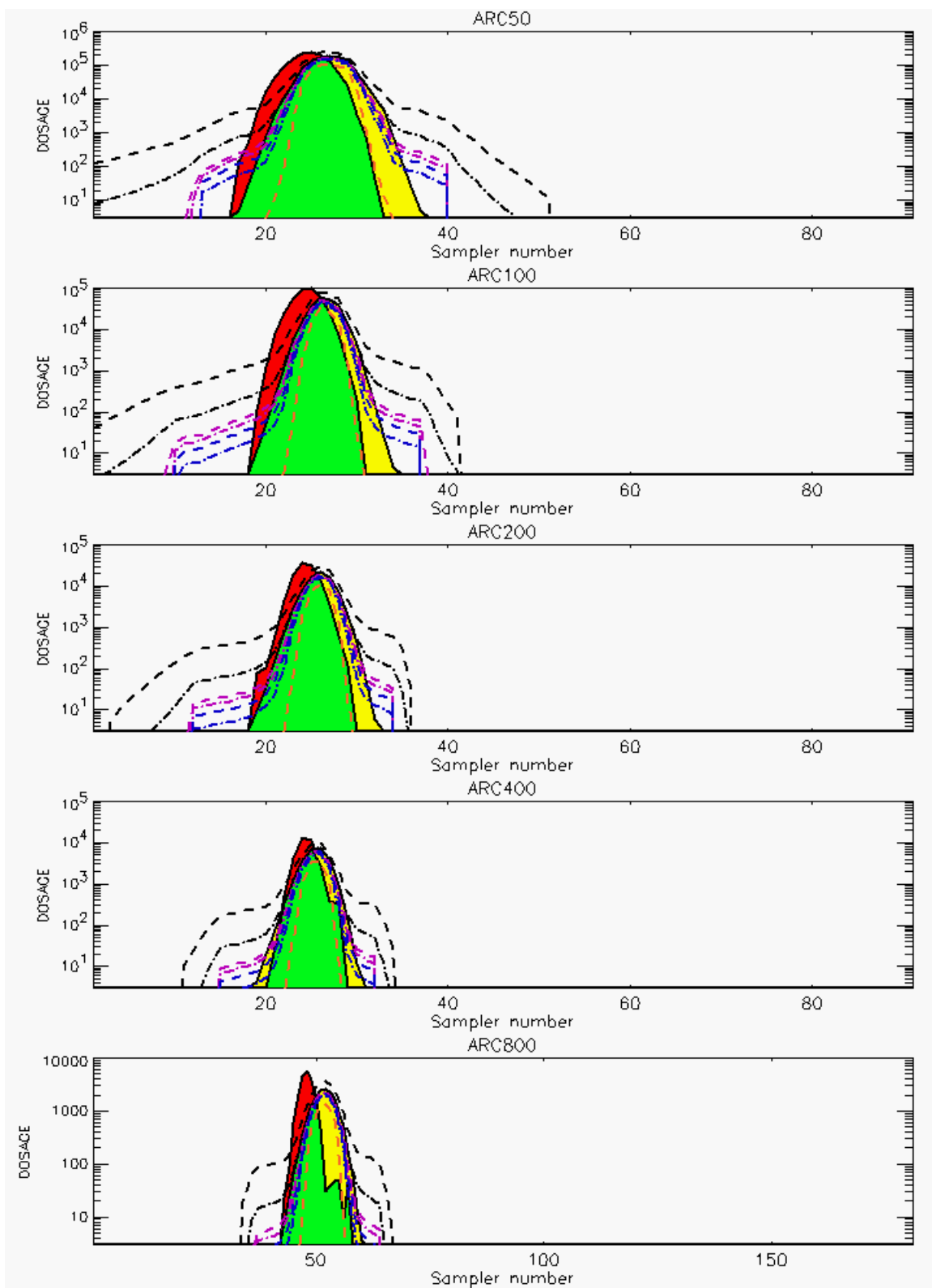


Figure C-43b. HPAC Probabilistic Prediction Outputs for Trial 54 on Logarithmic Scale: Stability Class is 5

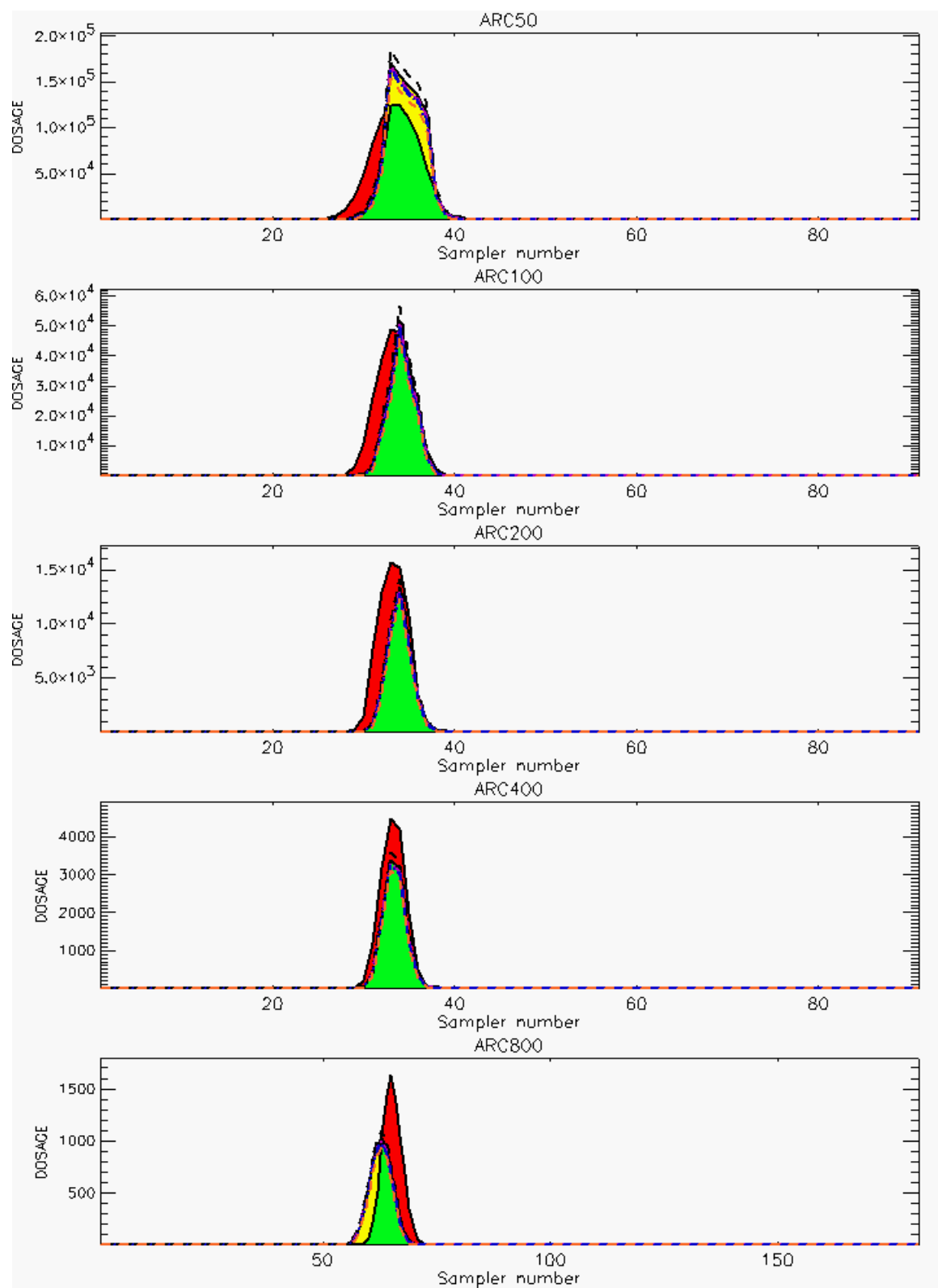
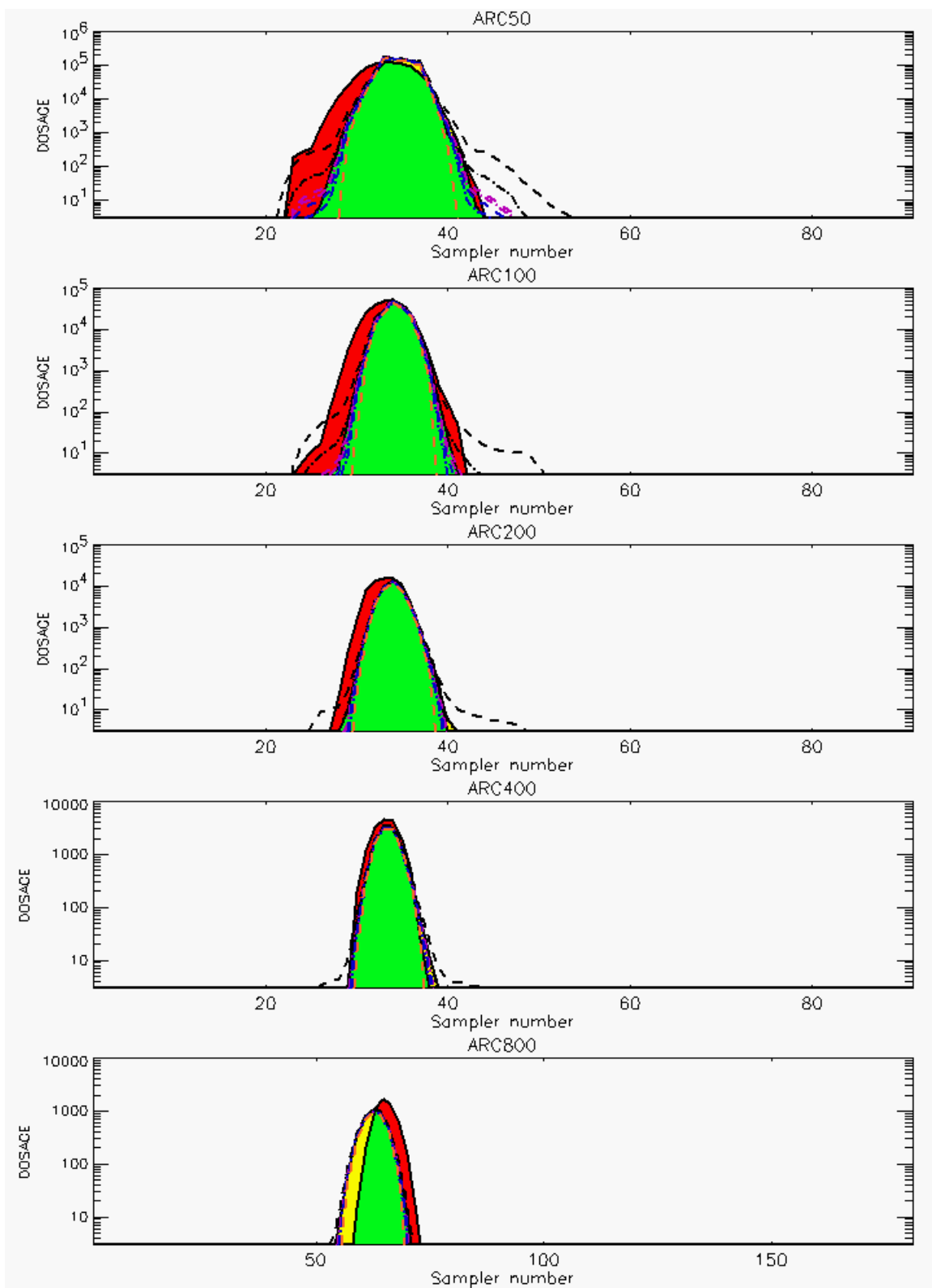


Figure C-44a. HPAC Probabilistic Prediction Outputs for Trial 55 on Linear Scale: Stability Class is 4



**Figure C-44b. HPAC Probabilistic Prediction Outputs for Trial 55 on Logarithmic Scale:
Stability Class is 4**

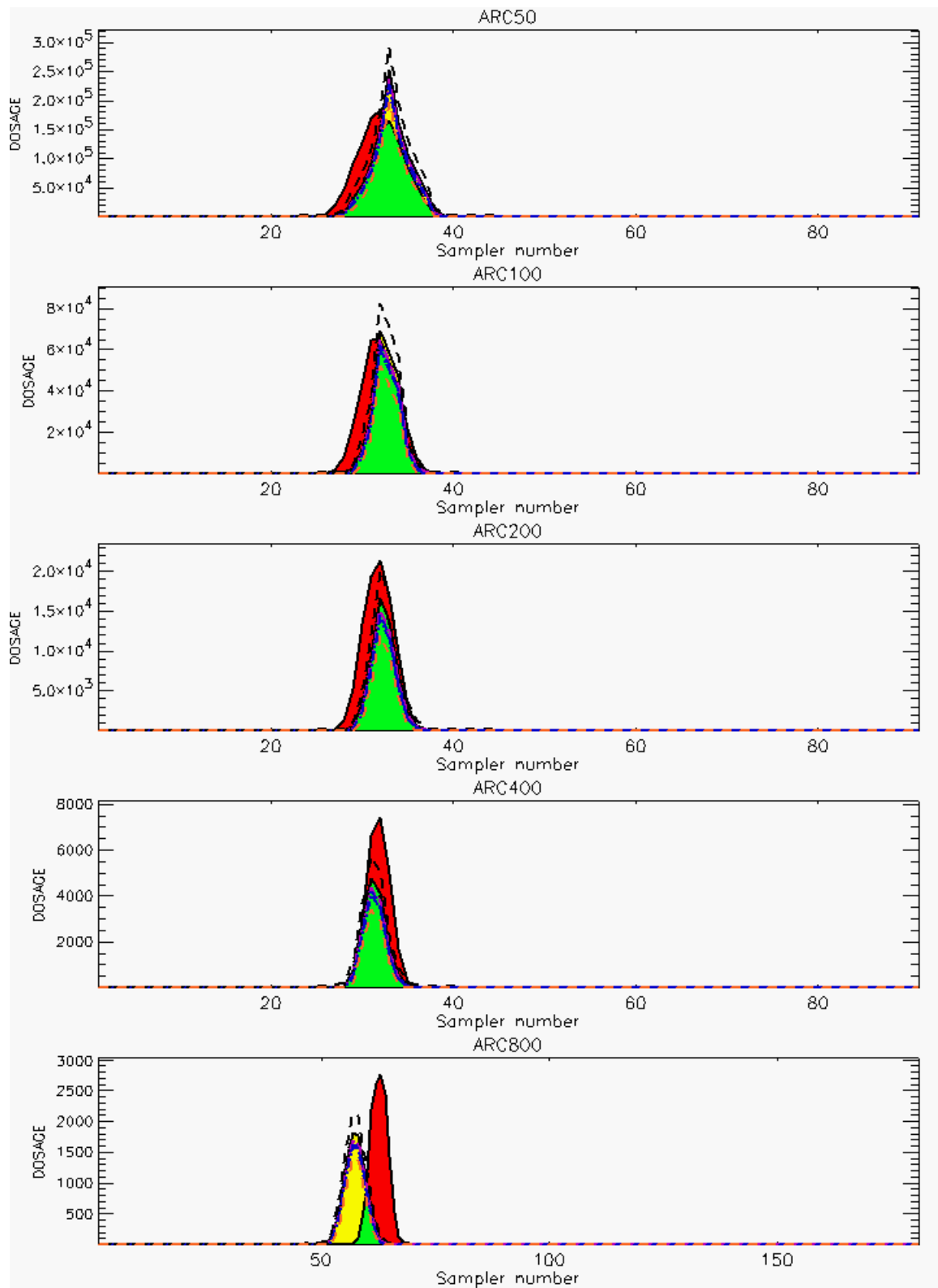


Figure C-45a. HPAC Probabilistic Prediction Outputs for Trial 56 on Linear Scale: Stability Class is 4

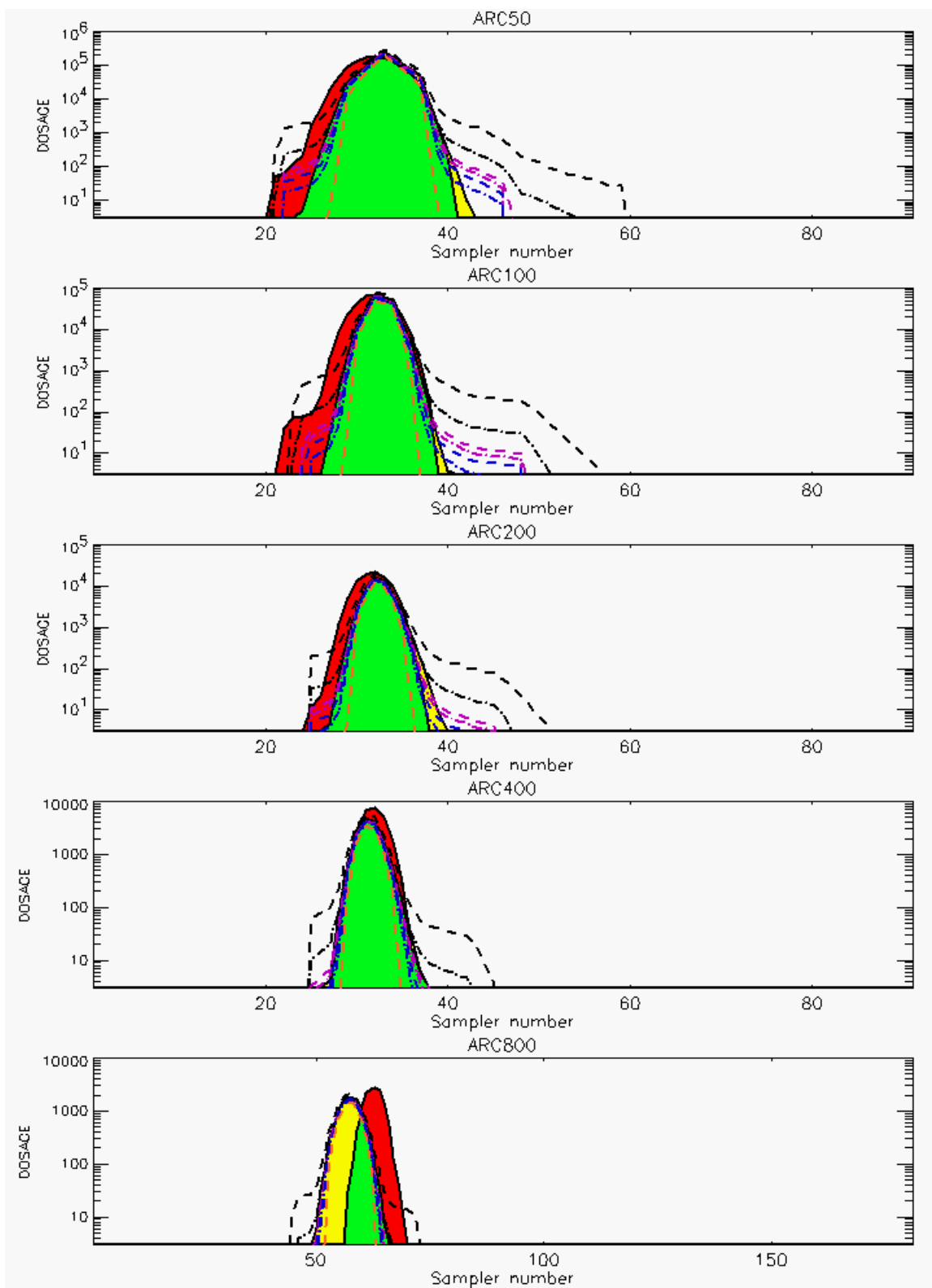


Figure C-45b. HPAC Probabilistic Prediction Outputs for Trial 56 on Logarithmic Scale: Stability Class is 4

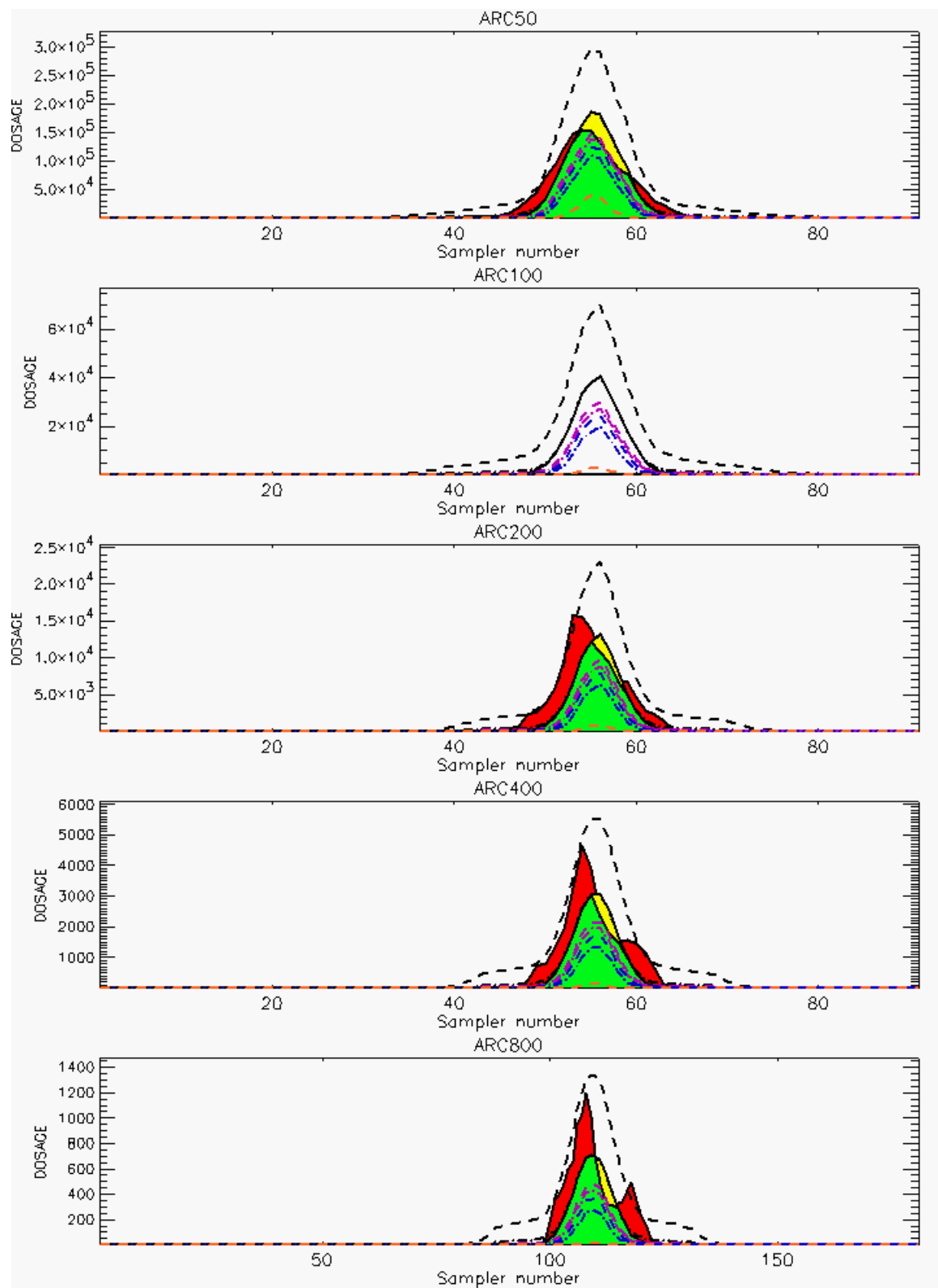


Figure C-46a. HPAC Probabilistic Prediction Outputs for Trial 57 on Linear Scale: Stability Class is 3 (Values for Sampler of 100-Meter Arc are Replaced by Missing Values [Ref. C-2])

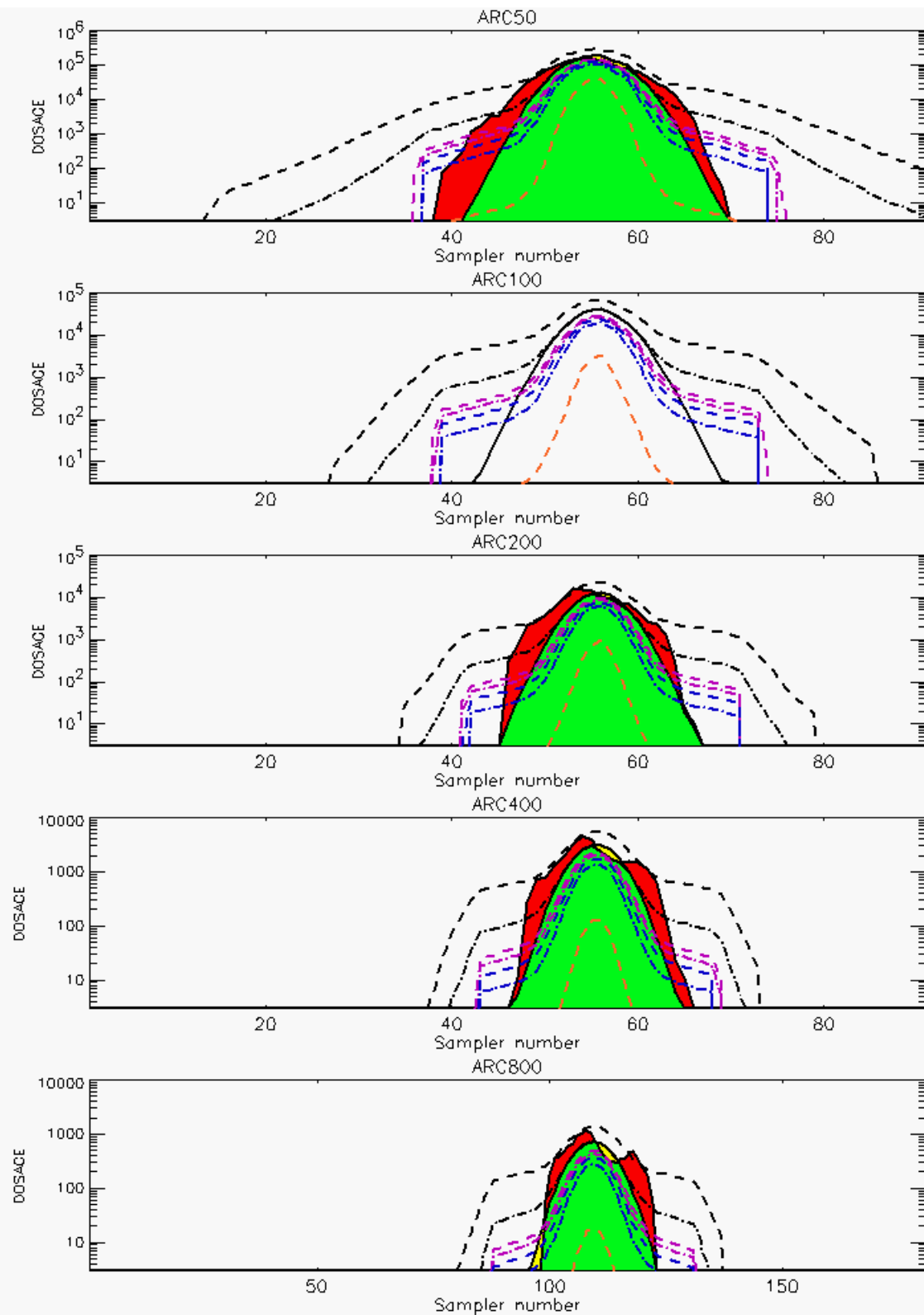


Figure C-46b. HPAC Probabilistic Prediction Outputs for Trial 57 on Logarithmic Scale: Stability Class is 3 (Values for Sampler of 100-Meter Arc are Replaced by Missing Values [Ref. C-2])

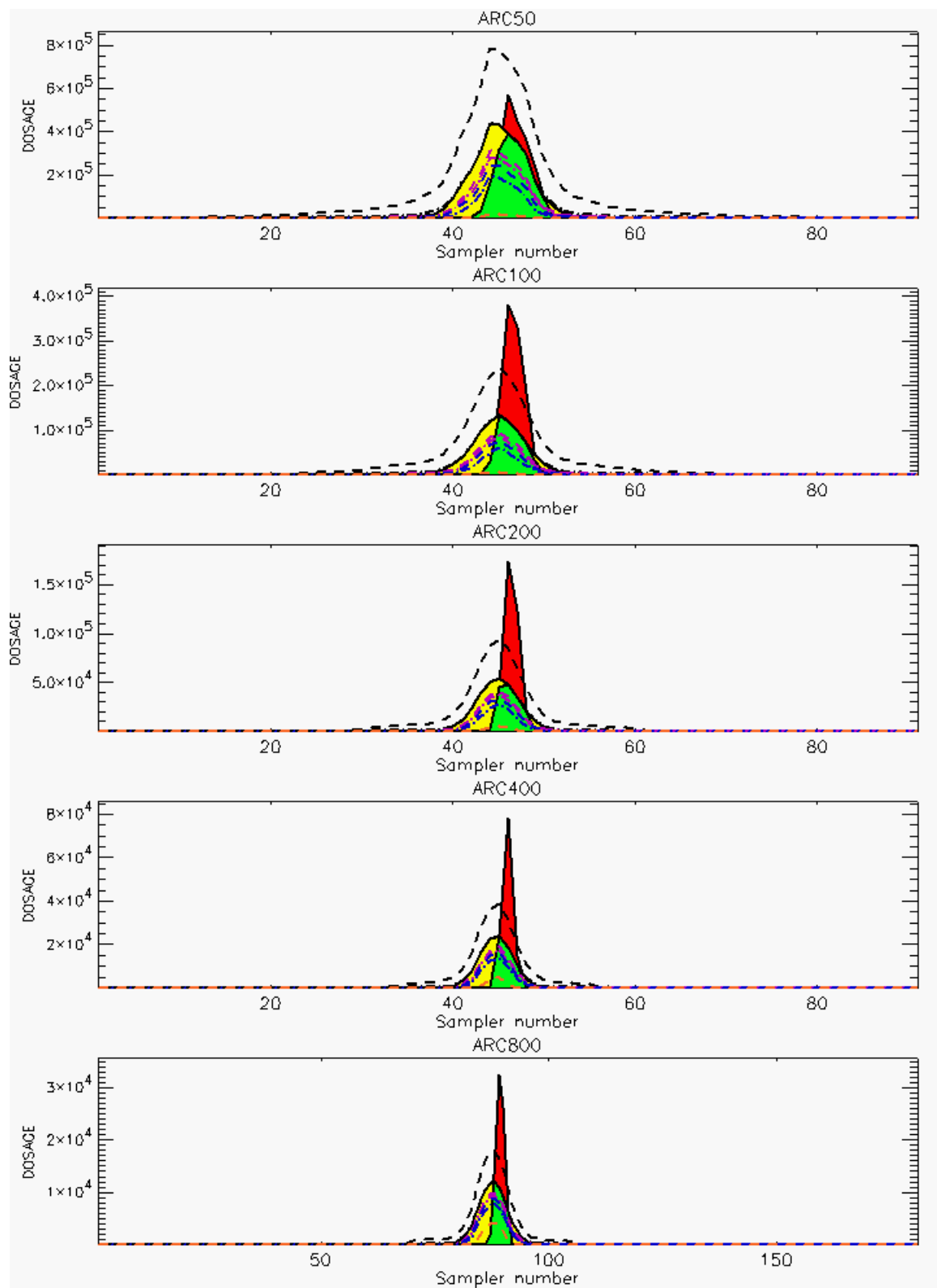


Figure C-47a. HPAC Probabilistic Prediction Outputs for Trial 58 on Linear Scale: Stability Class is 6

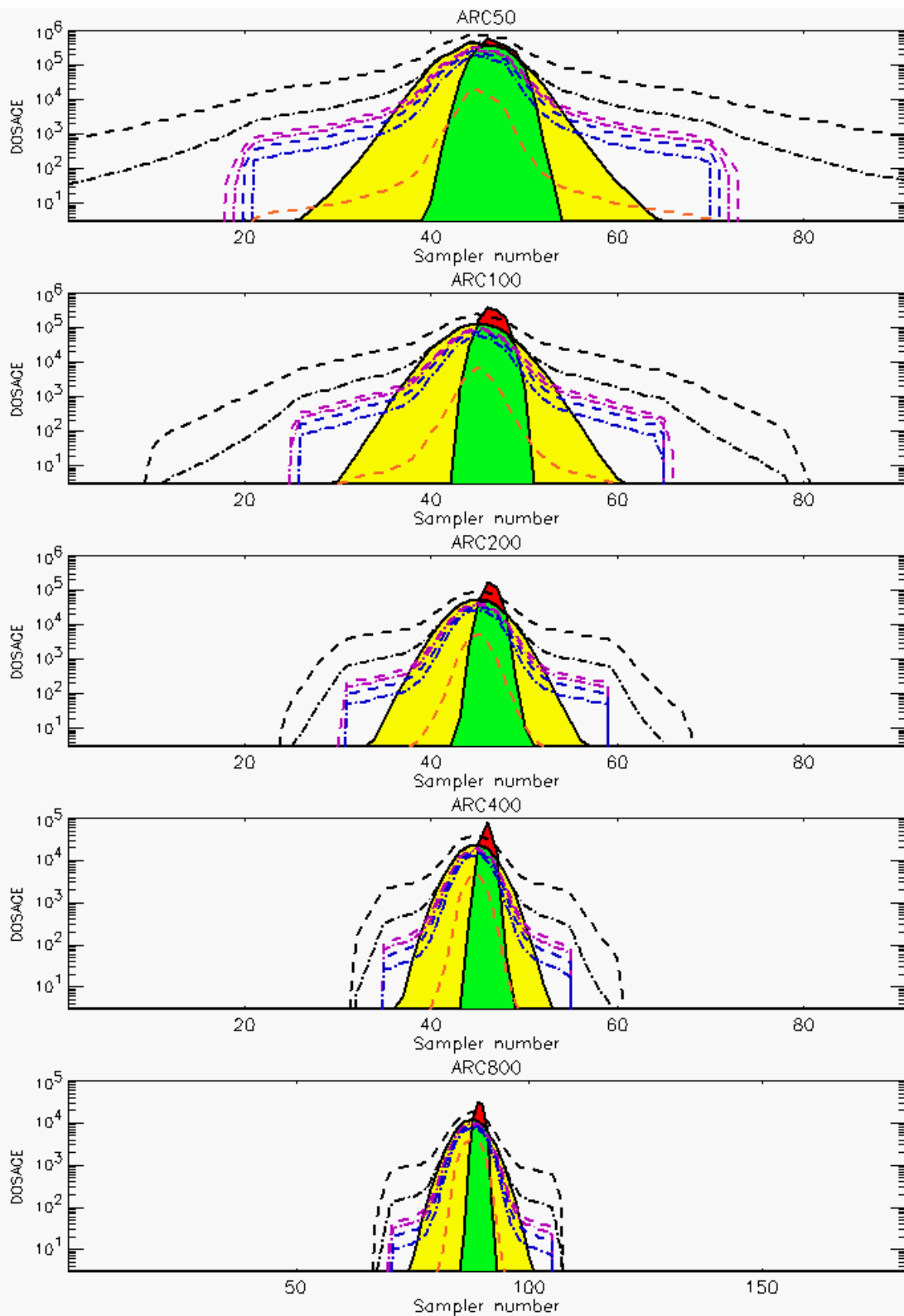


Figure C-47b. HPAC Probabilistic Prediction Outputs for Trial 58 on Logarithmic Scale: Stability Class is 6

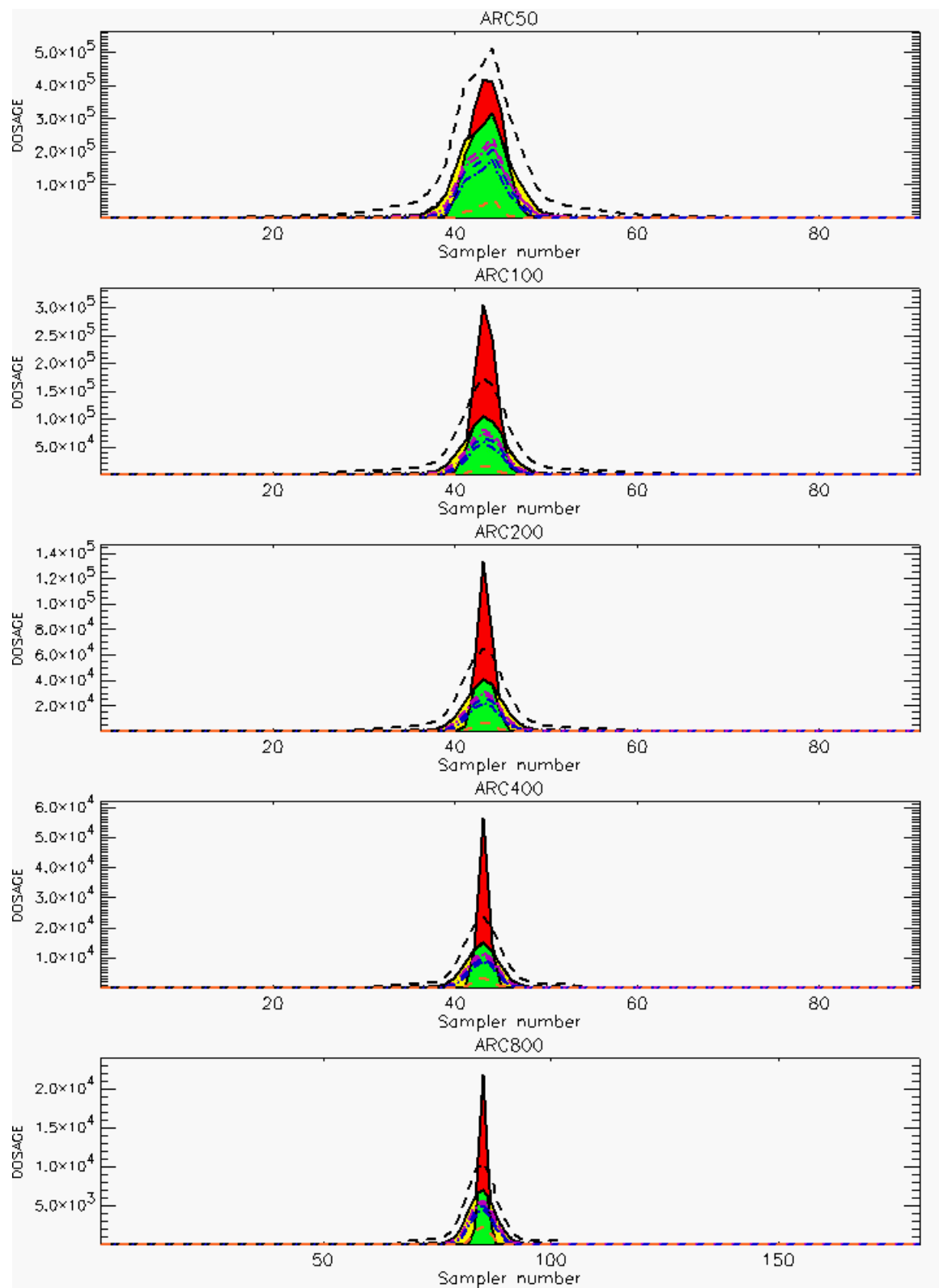
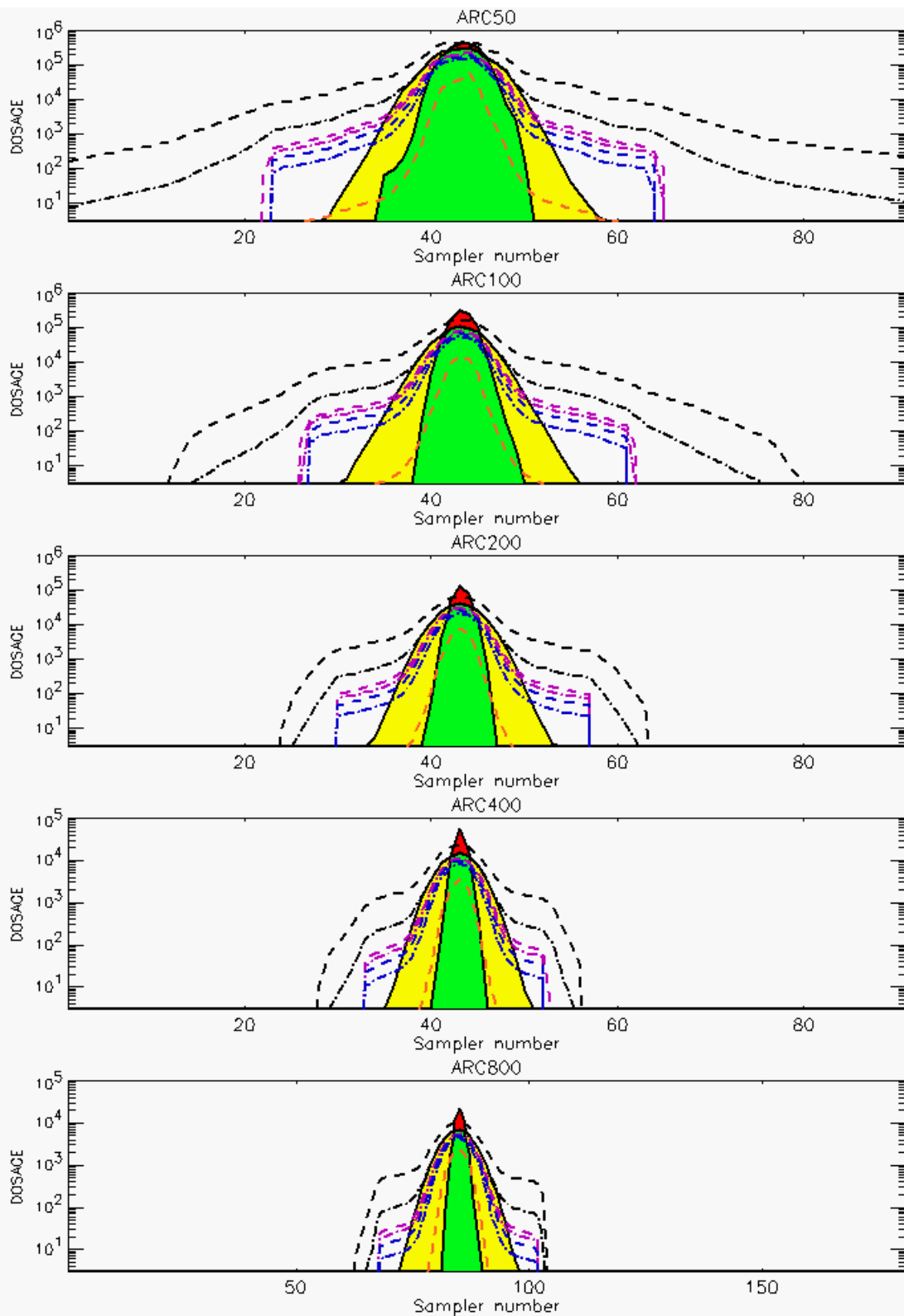


Figure C-48a. HPAC Probabilistic Prediction Outputs for Trial 59 on Linear Scale: Stability Class is 5



**Figure C-48b. HPAC Probabilistic Prediction Outputs for Trial 59 on Logarithmic Scale:
Stability Class is 5**

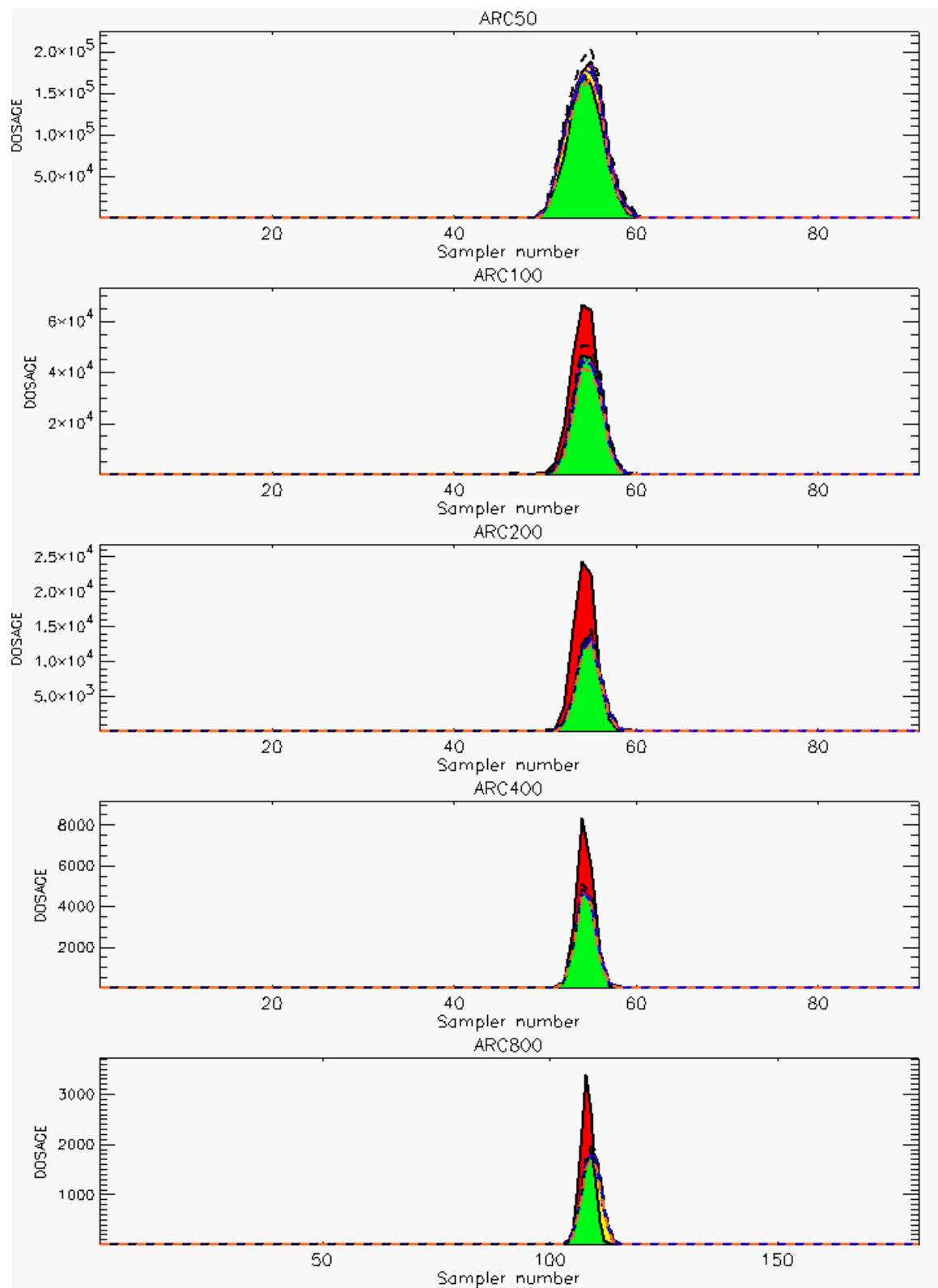


Figure C-49a. HPAC Probabilistic Prediction Outputs for Trial 60 on Linear Scale: Stability Class is 5

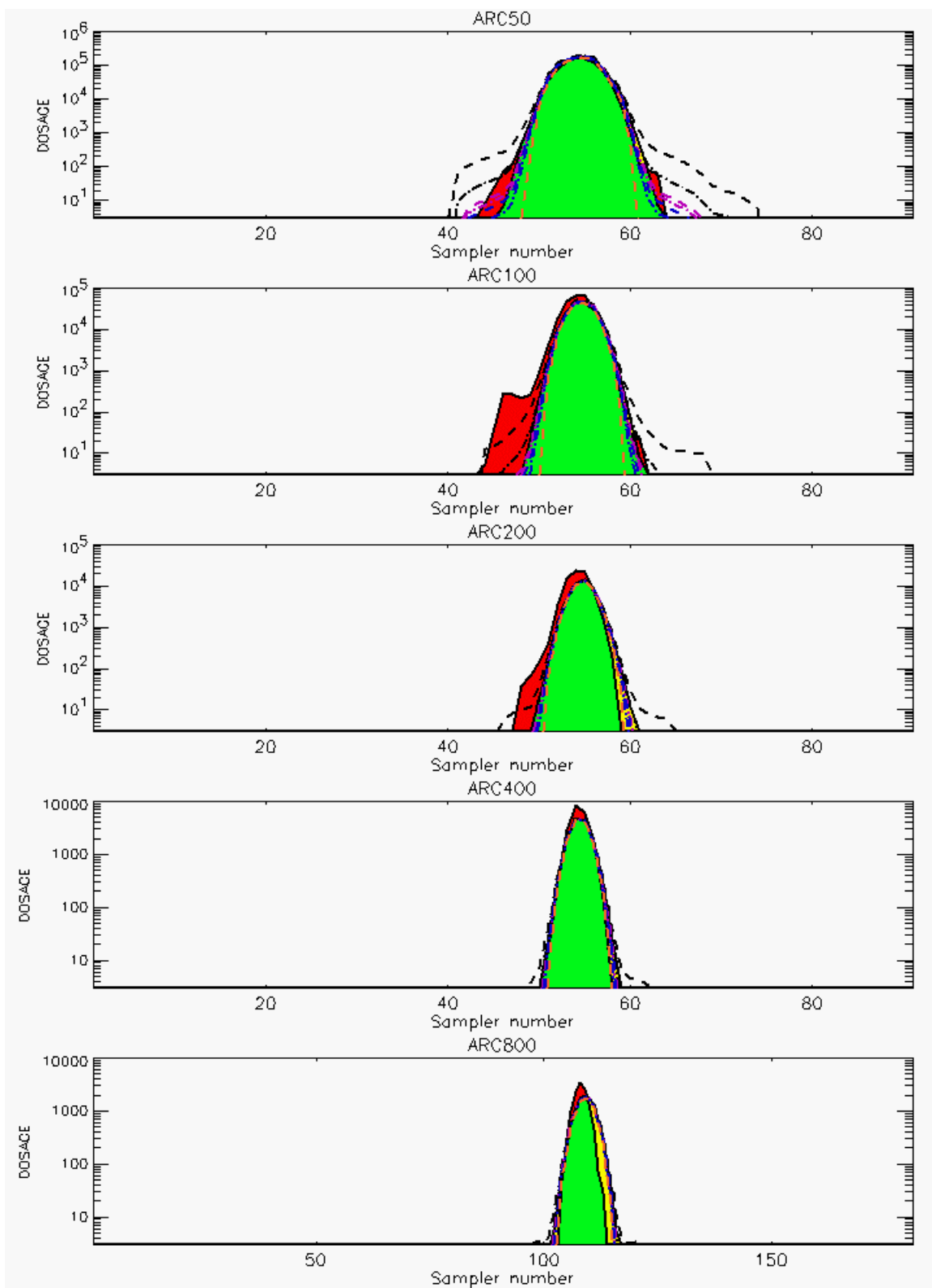


Figure C-49b. HPAC Probabilistic Prediction Outputs for Trial 60 on Logarithmic Scale: Stability Class is 5

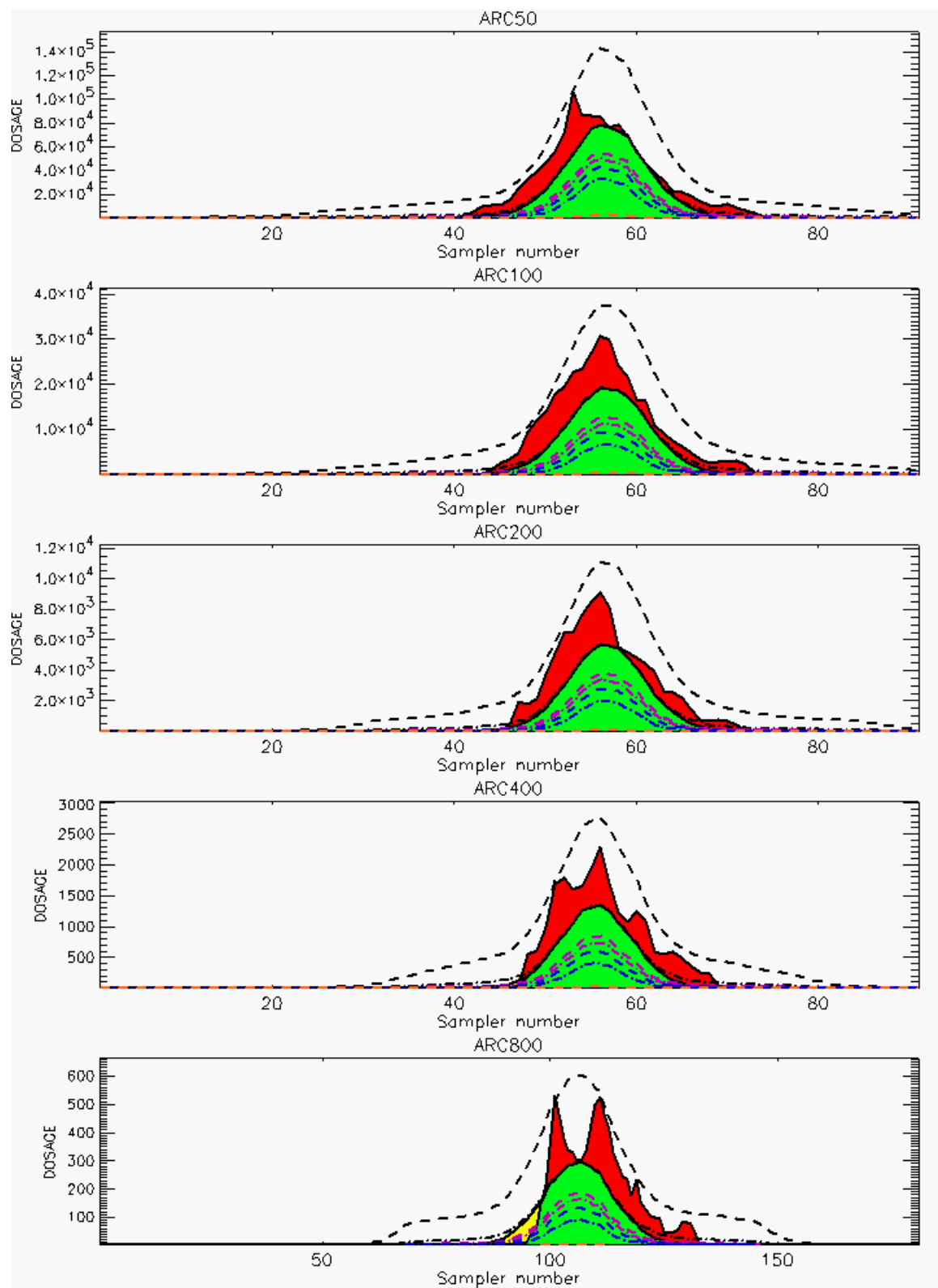


Figure C-50a. HPAC Probabilistic Prediction Outputs for Trial 61 on Linear Scale: Stability Class is 3

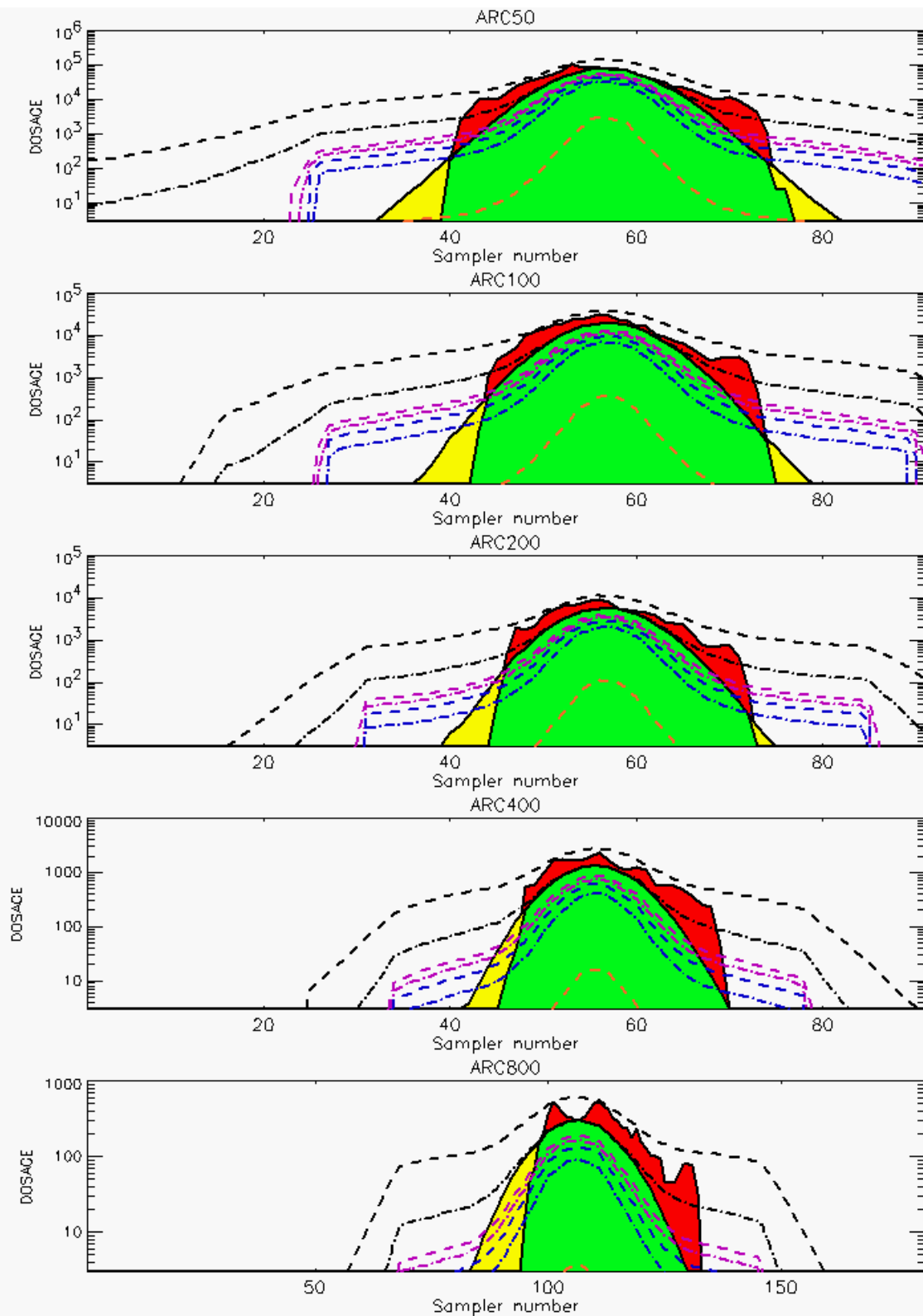


Figure C-50b. HPAC Probabilistic Prediction Outputs for Trial 61 on Logarithmic Scale: Stability Class is 3

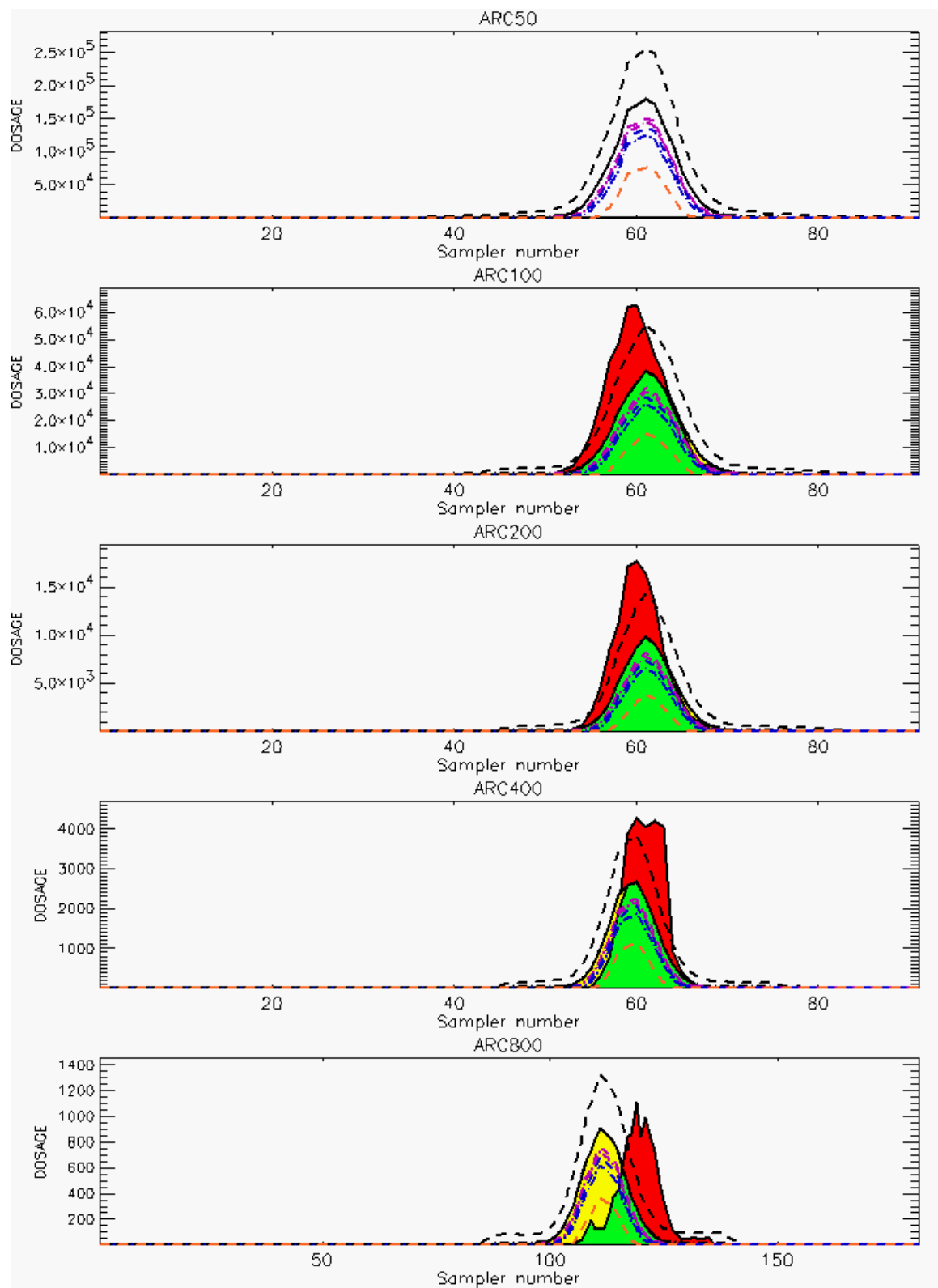


Figure C-51a. HPAC Probabilistic Prediction Outputs for Trial 62 on Linear Scale: Stability Class is 2 (Values for Samplers of 50-Meter Arc are Replaced by Missing Values [Ref. C-2])

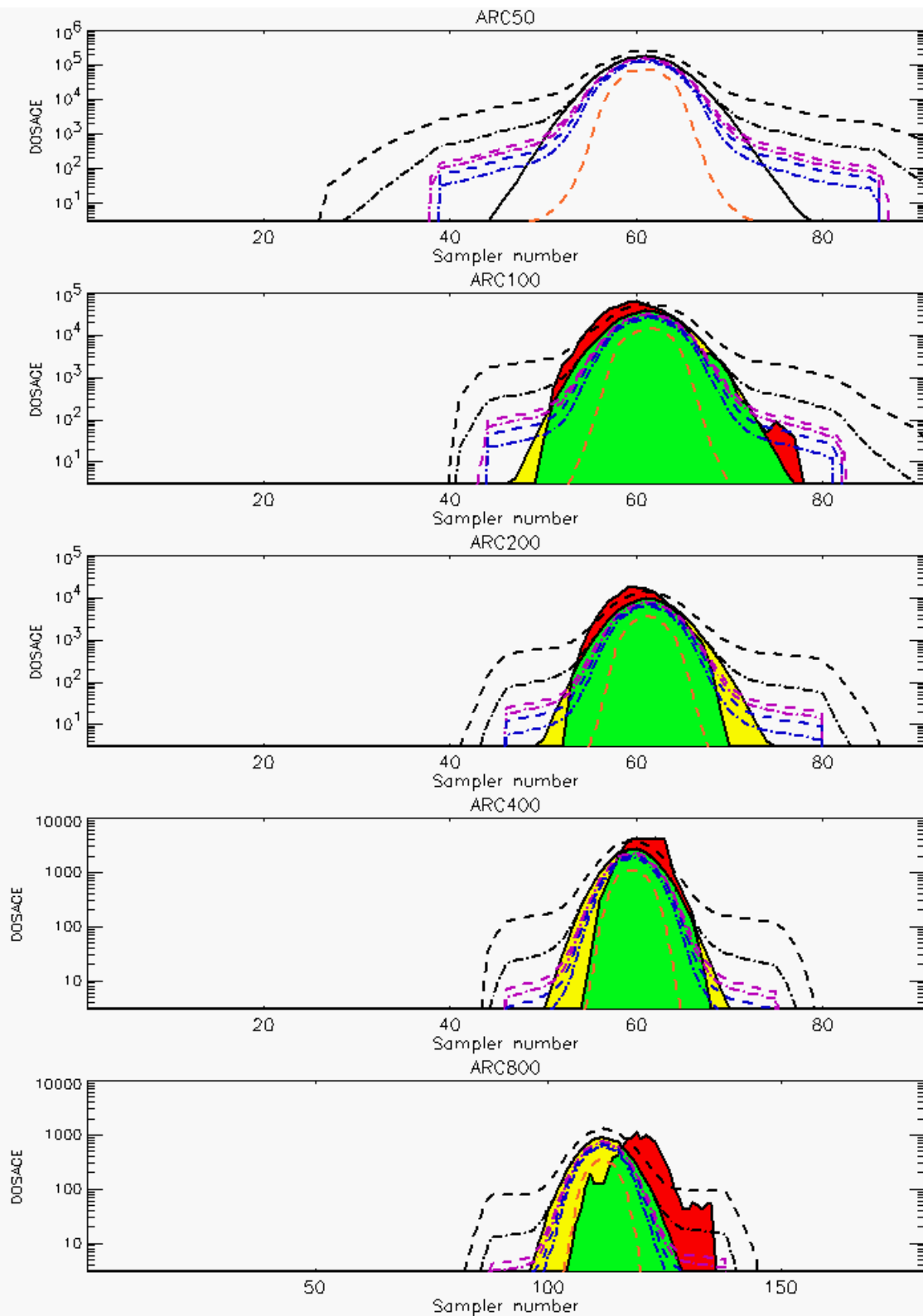


Figure C-51b. HPAC Probabilistic Prediction Outputs for Trial 62 on Logarithmic Scale: Stability Class is 2 (Values for Samplers of 50-Meter Arc are Replaced by Missing Values [Ref. C-2])

REFERENCES

- C-1. Barad, M. L. (Editor), *Project Prairie Grass, A Field Program in Diffusion*, Geophysical Research Papers No. 59, Volumes I and II, DTIC #AD-152572/AFCRC-TR-58-235(I), Air Geophysical Laboratory, Hanscom Air Force Base, MA, 1958.
- C-2. Barad, M. L. (Editor), *Project Prairie Grass, a Field Program in Diffusion*, Geophysical Research Papers, No. 59, Volume I, DTIC #AD-152572/AFCRC-TR-58-235(I), pages 79-80, 201, 1958.
- C-3. Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T, and Larson, D., *User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models*, IDA Paper P-3554, Appendix K, January 2001.
- C-4. Irwin, J. S. and Rosu, M-R., "Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models," *Proceedings of the 10th Joint Conference on the Applications of Air Pollution Meteorology*, American Meteorological Society, Boston, pp. 6-10, 1998.

APPENDIX D
HPAC PROBABILISTIC OUTPUTS: FULL VERSUS
CONDITIONAL PROBABILITIES FOR *PRAIRIE GRASS*
PREDICTIONS

APPENDIX D

HPAC PROBABILISTIC OUTPUTS: FULL (0.50) VERSUS CONDITIONAL (0.50) PROBABILITIES FOR *PRAIRIE GRASS* PREDICTIONS

This appendix presents graphical comparison of HPAC probabilistic value prediction outputs to *Prairie Grass* field trials [Ref. D-1].¹ Vertical plot units are dosage units of mg-sec/m³. Horizontal plot units are sampler numbers as presented in the *Prairie Grass* field trials with sampler number 1 oriented to the west, the middle sampler (45 or 90) oriented to the north, and the last sampler (91 or 181) oriented to the east of the SO₂ gas release source. Only data values greater than the cutoff threshold of 3 mg-sec/m³ (i.e., an average concentration of 0.005 mg/m³ over 600 seconds) are presented for both field trial data and HPAC probabilistic prediction outputs. This cutoff threshold value corresponds to a minimum value reported in *Prairie Grass* field trials.

Comparisons of HPAC probabilistic and mean value prediction outputs, as well as *Prairie Grass* field trial observations, are presented on both linear and logarithmic dosage scales. Each graphical comparison consists of five plots (one for each arc) with the top plot depicting the 50-meter arc, the second plot depicting the 100-meter arc, the third plot depicting the 200-meter arc, and so on. All of the figures of this appendix are presented on a logarithmic dosage scale.

The solid lines correspond to the observed values and the 0.50 conditional probability prediction (denoted “Exceed. Value (V[P_c>P])” on the HPAC plot screen in the *Plot Choice* dialog box). The meanings of the colors used in the plots are described in the below legend:

- Red — trial dosages that are higher than the 0.50 conditional probability prediction
- Green — trial dosages that overlap with the 0.50 conditional probability prediction

¹ These predictions were done without the inclusion of an SO₂ surface deposition mechanism.

- Yellow — 0.50 conditional probability predicted dosages that are higher than the trial observations
- White — trial sampler observation is missing.

The blue diamonds denote samplers where HPAC predicted a 0.50 (or greater) probability of exceeding 60 mg-sec/m³. These outputs are denoted as “Probability (P [V>E])” on the HPAC plot screen in the *Plot Choice* dialog box. The 60 mg-sec/m³ dosage level is highlighted by dashed lines in the below figures.

Samplers with missing values, or spurious maximum that were fixed by moving decimal point in the analyses, are noted in the figure captions [Ref. D-2]. Additional information is contained in Reference D-3.

Irwin stability categories that were used in our analyses are denoted in the figure captions. These stability category assignments are based on Reference D-4.

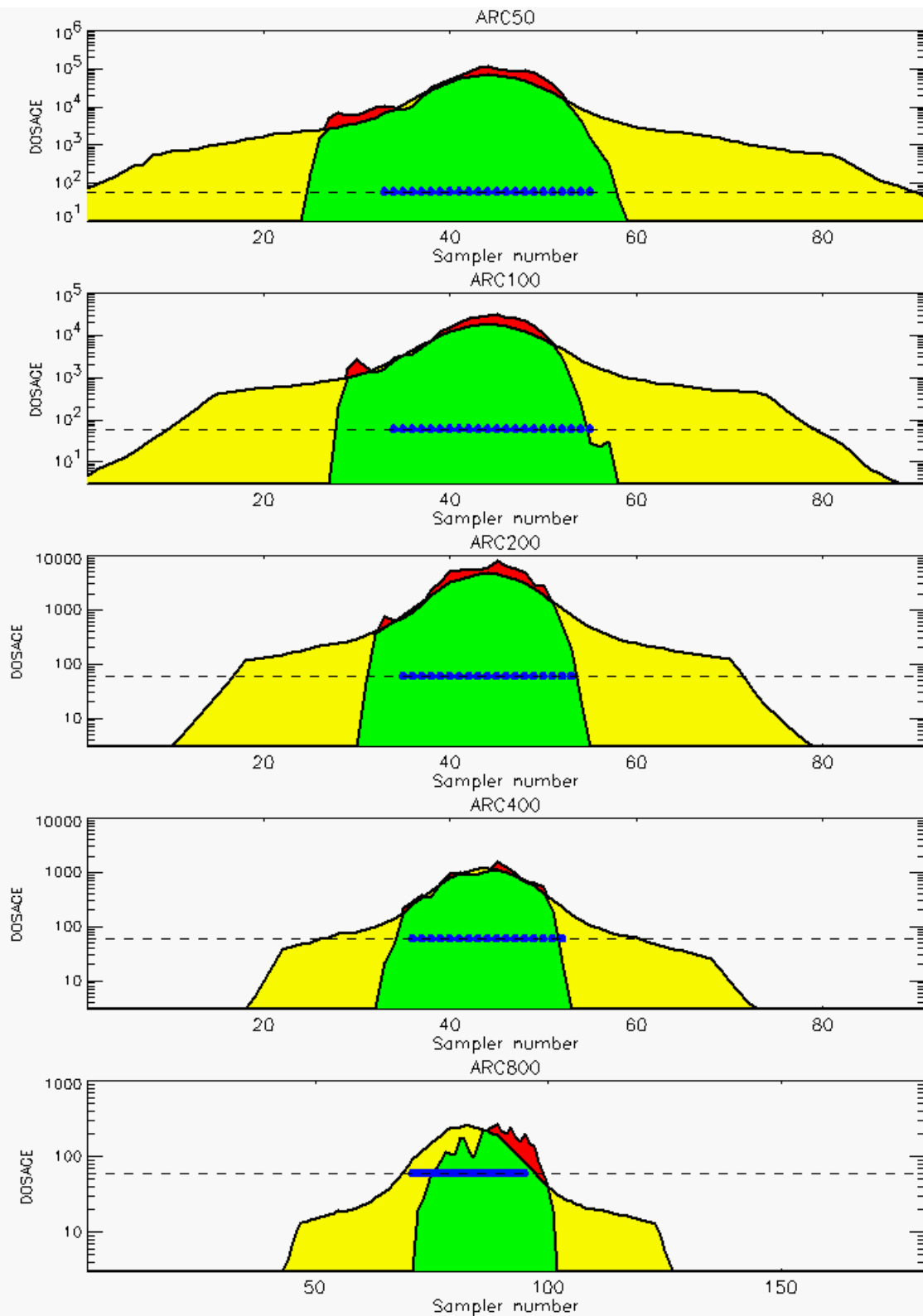


Figure D-1. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 5: Stability Class is 2

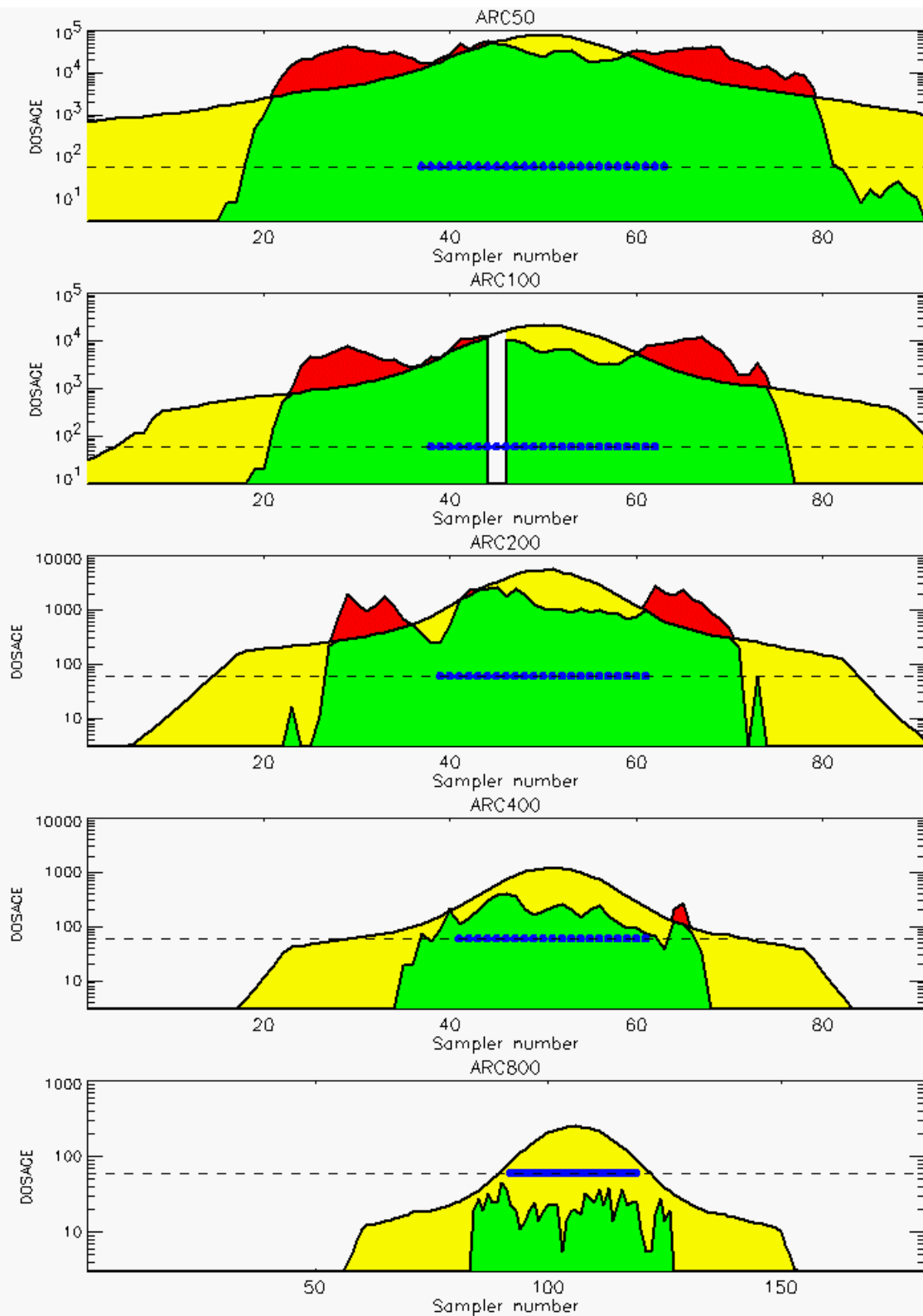


Figure D-2. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 7: Stability Class is 1

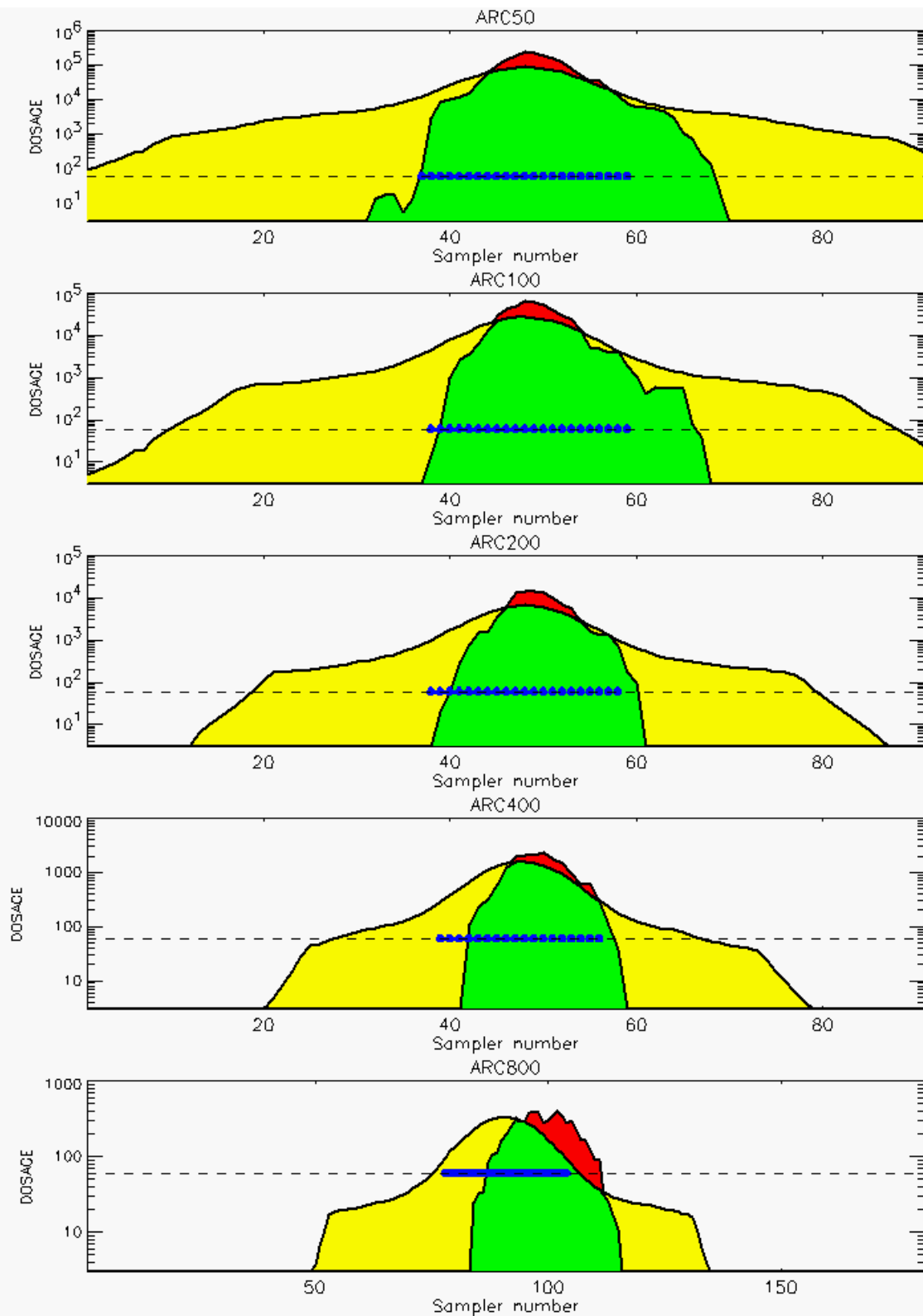


Figure D-3. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 8: Stability Class is 2

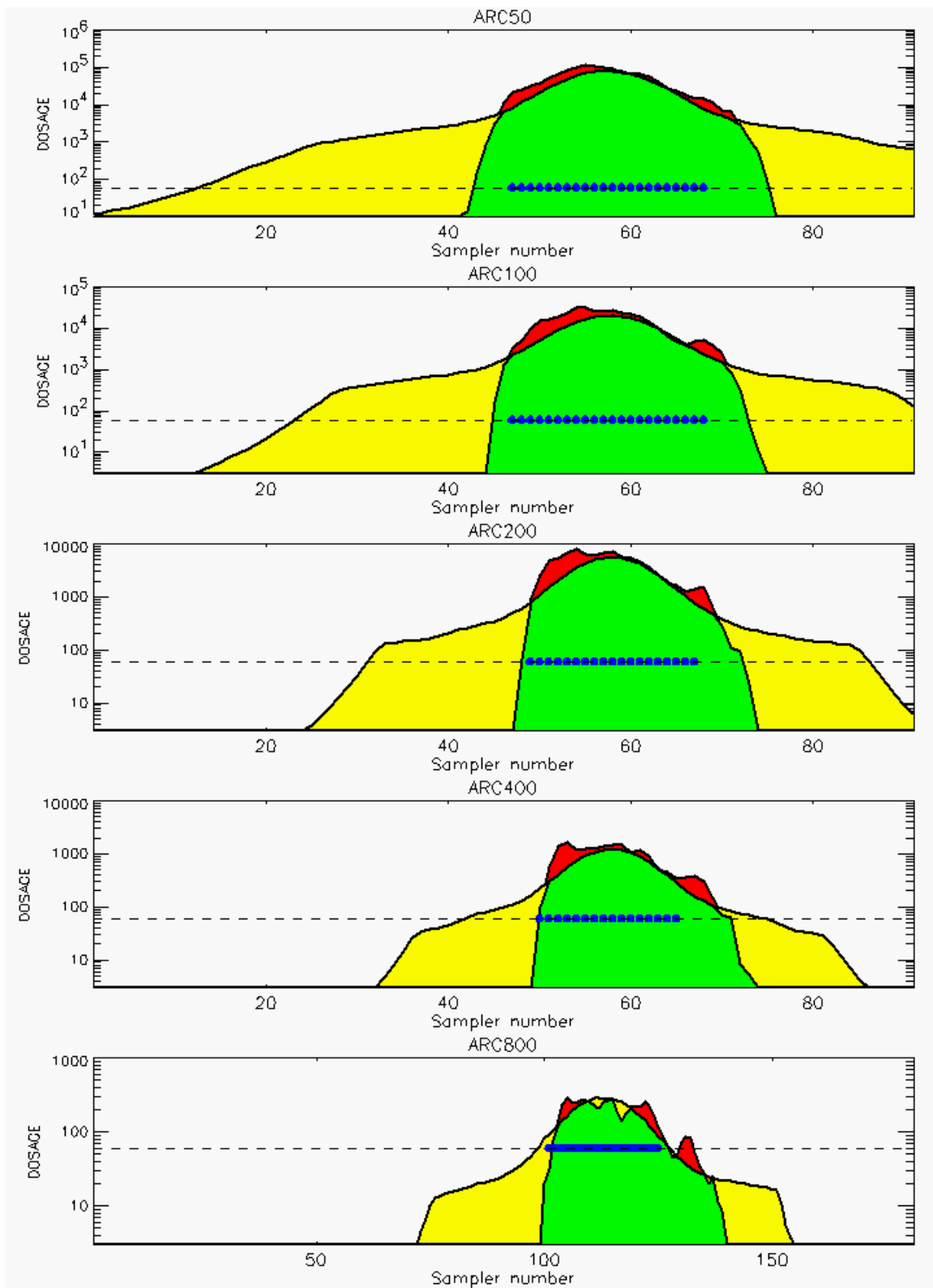


Figure D-4. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 9: Stability Class is 2

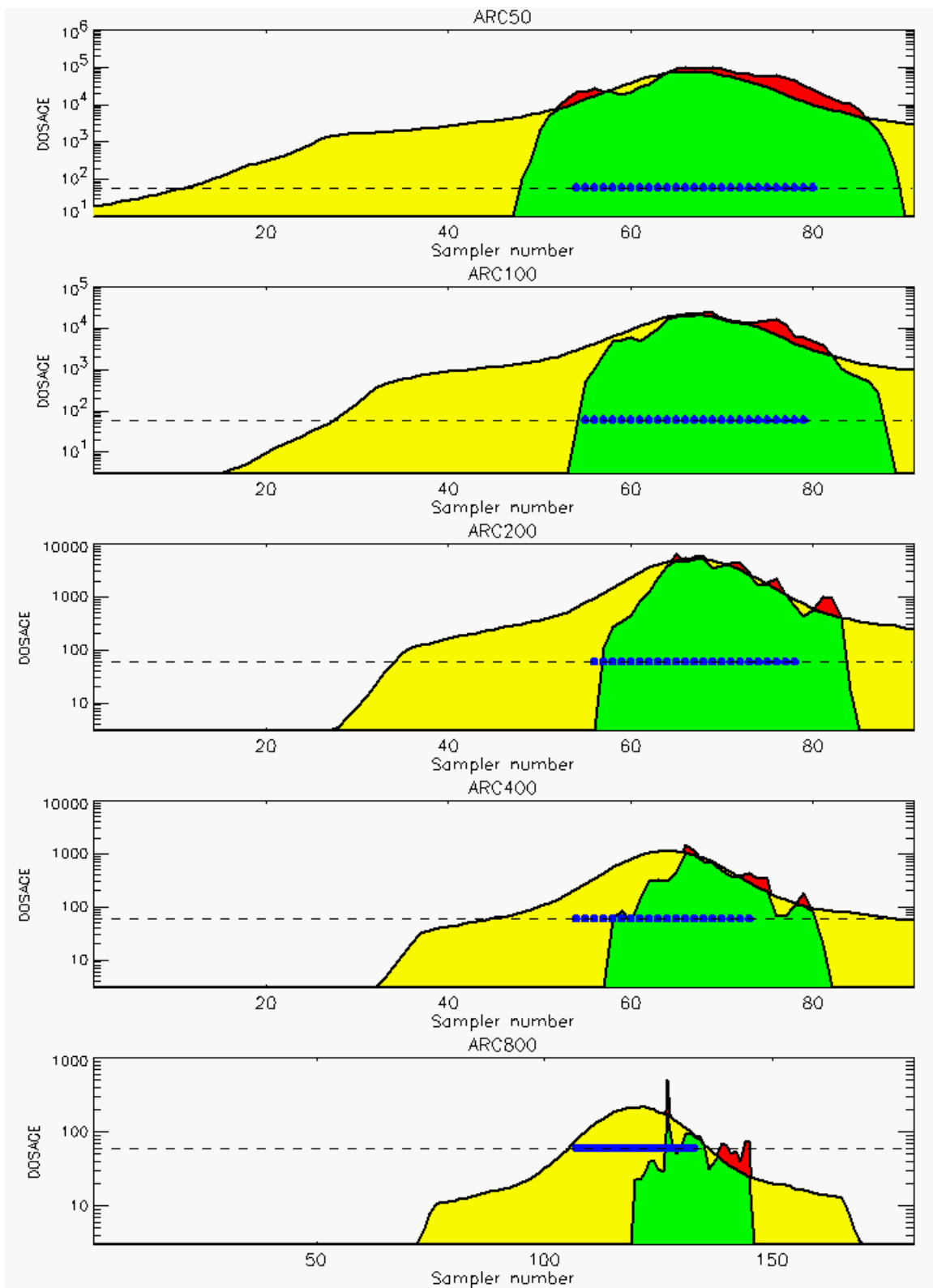


Figure D-5. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 10: Stability Class is 1

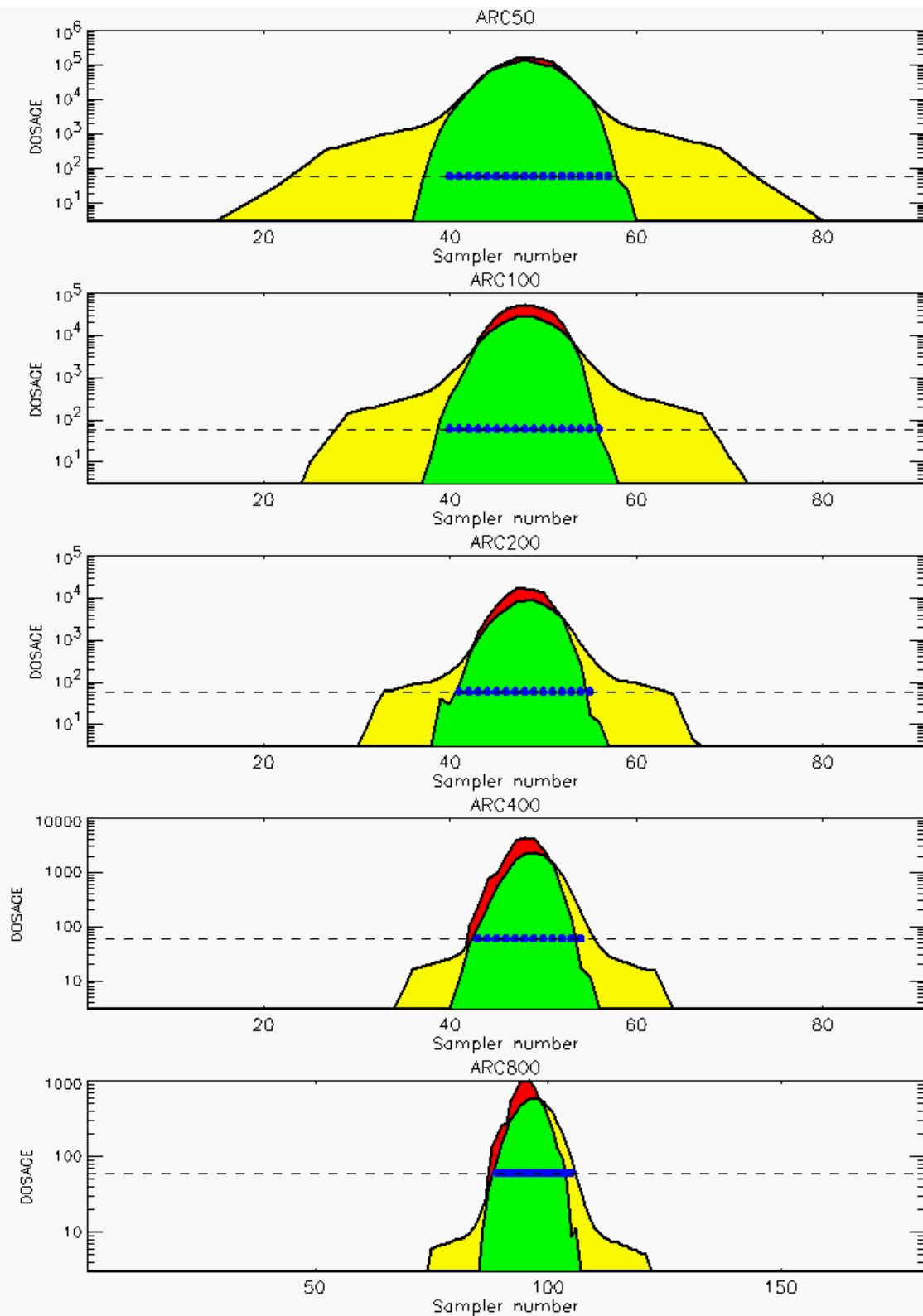


Figure D-6. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 11: Stability Class is 3

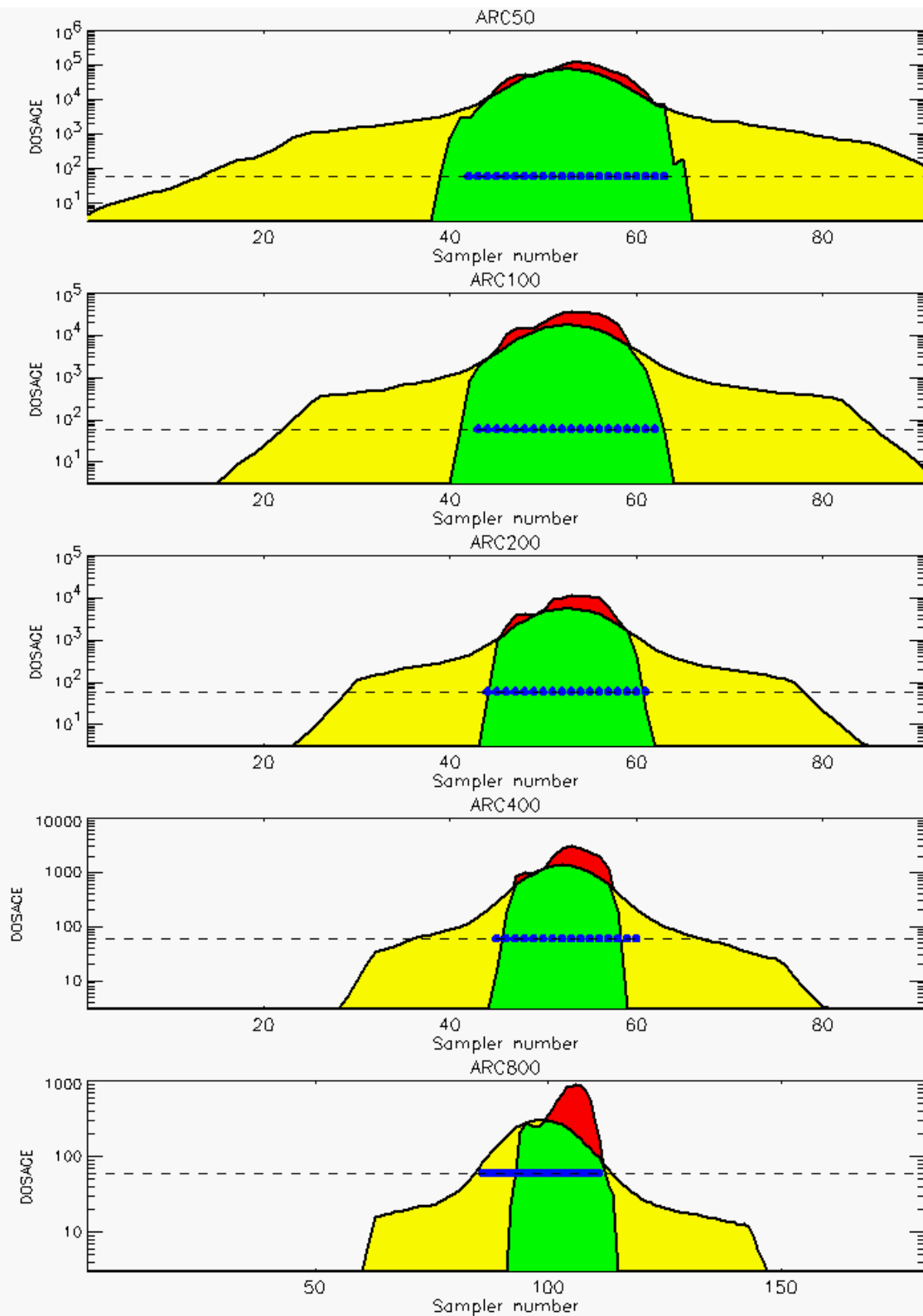


Figure D-7. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 12: Stability Class is 3

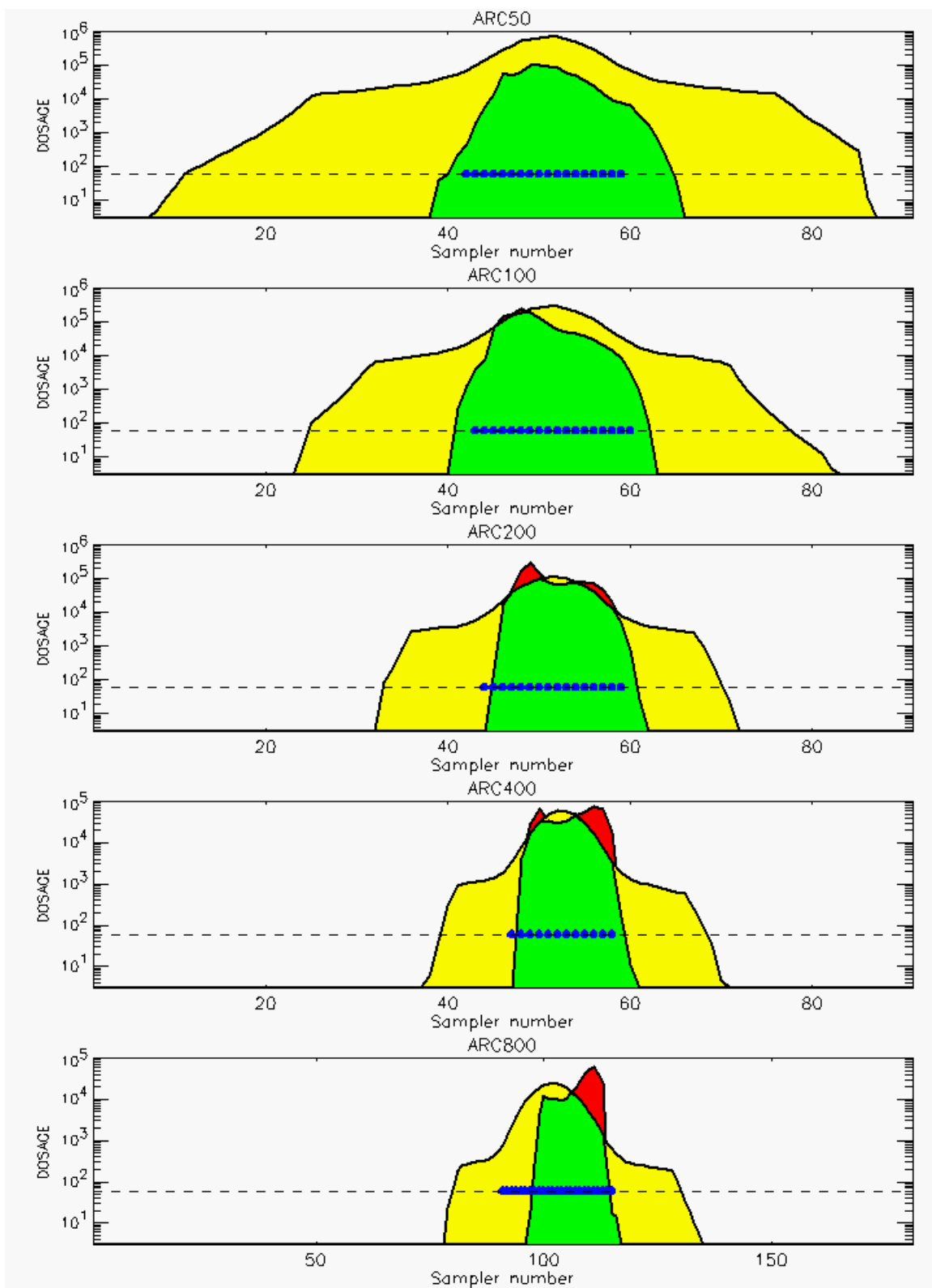


Figure D-8. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 13: Stability Class is 7

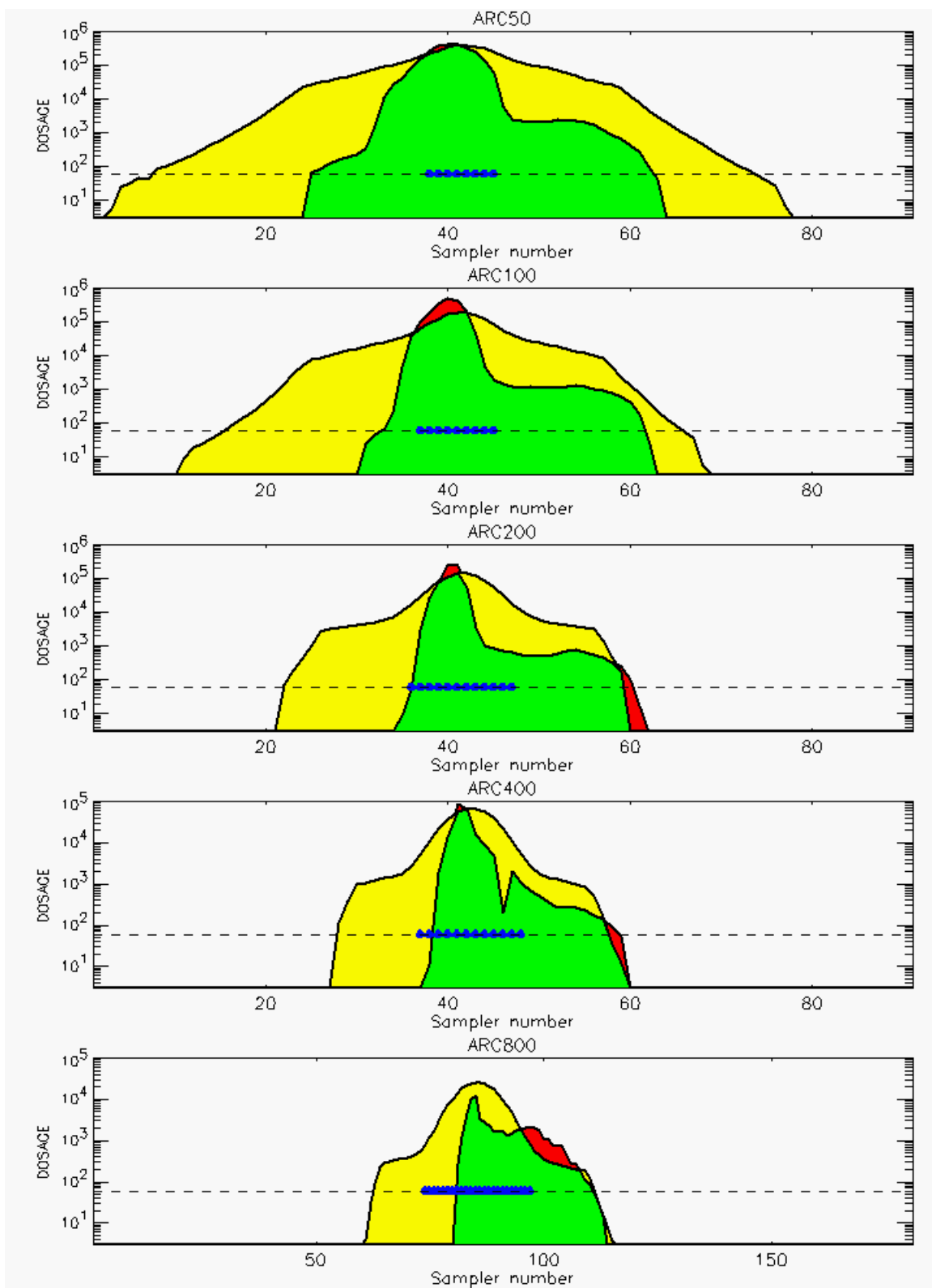


Figure D-9. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 14: Stability Class is 7

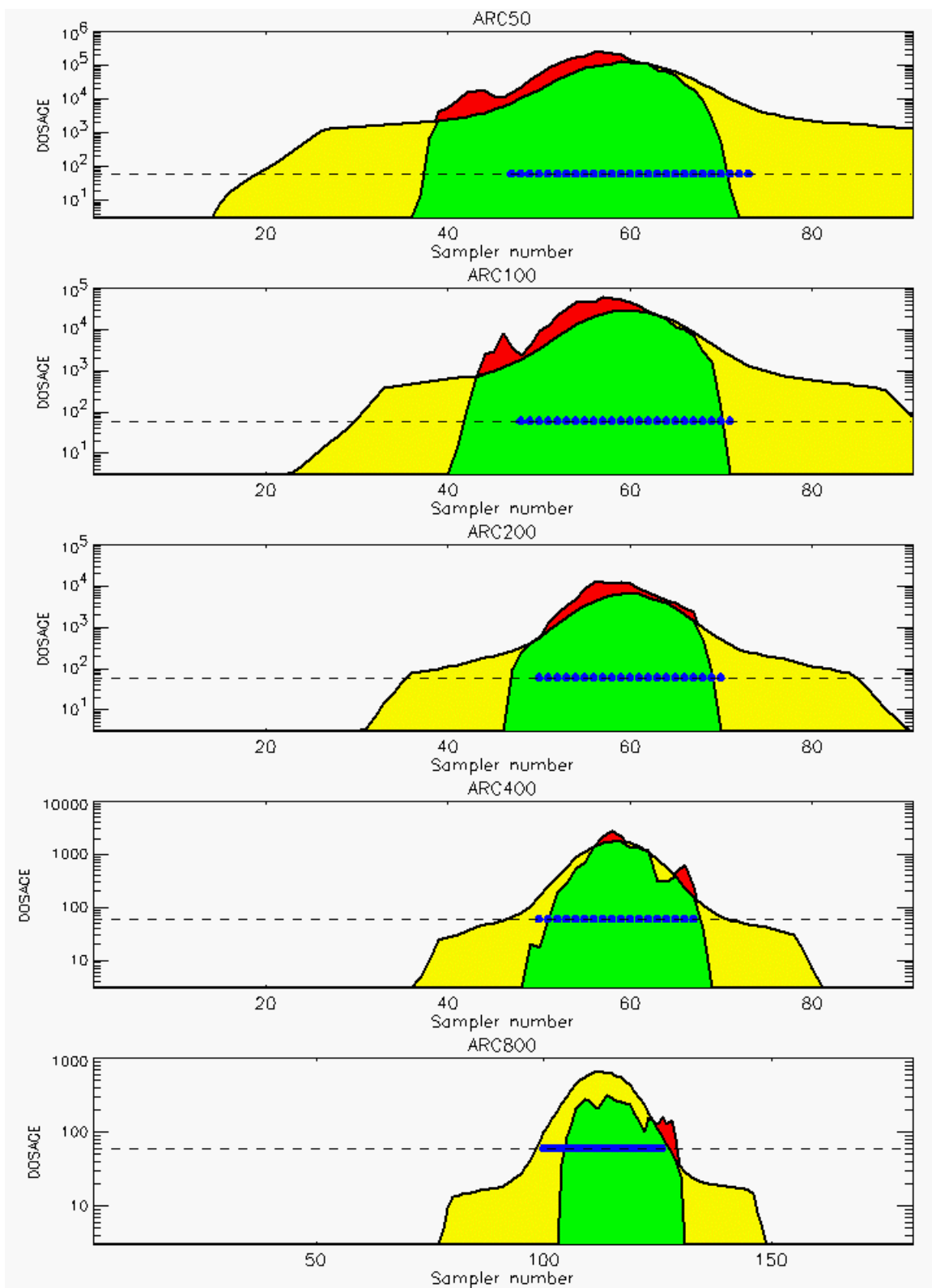


Figure D-10. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 15: Stability Class is 1

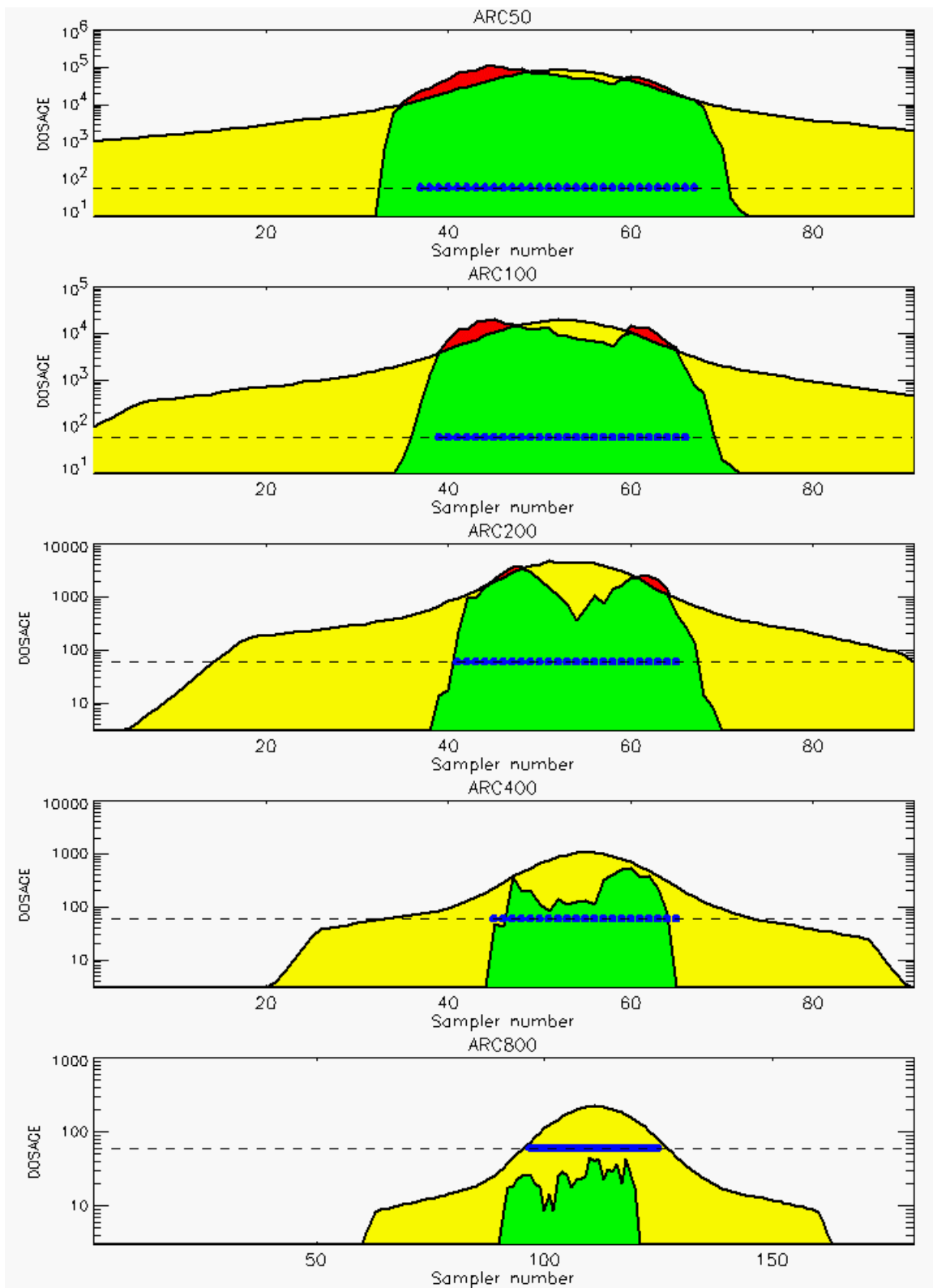


Figure D-11. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 16: Stability Class is 1

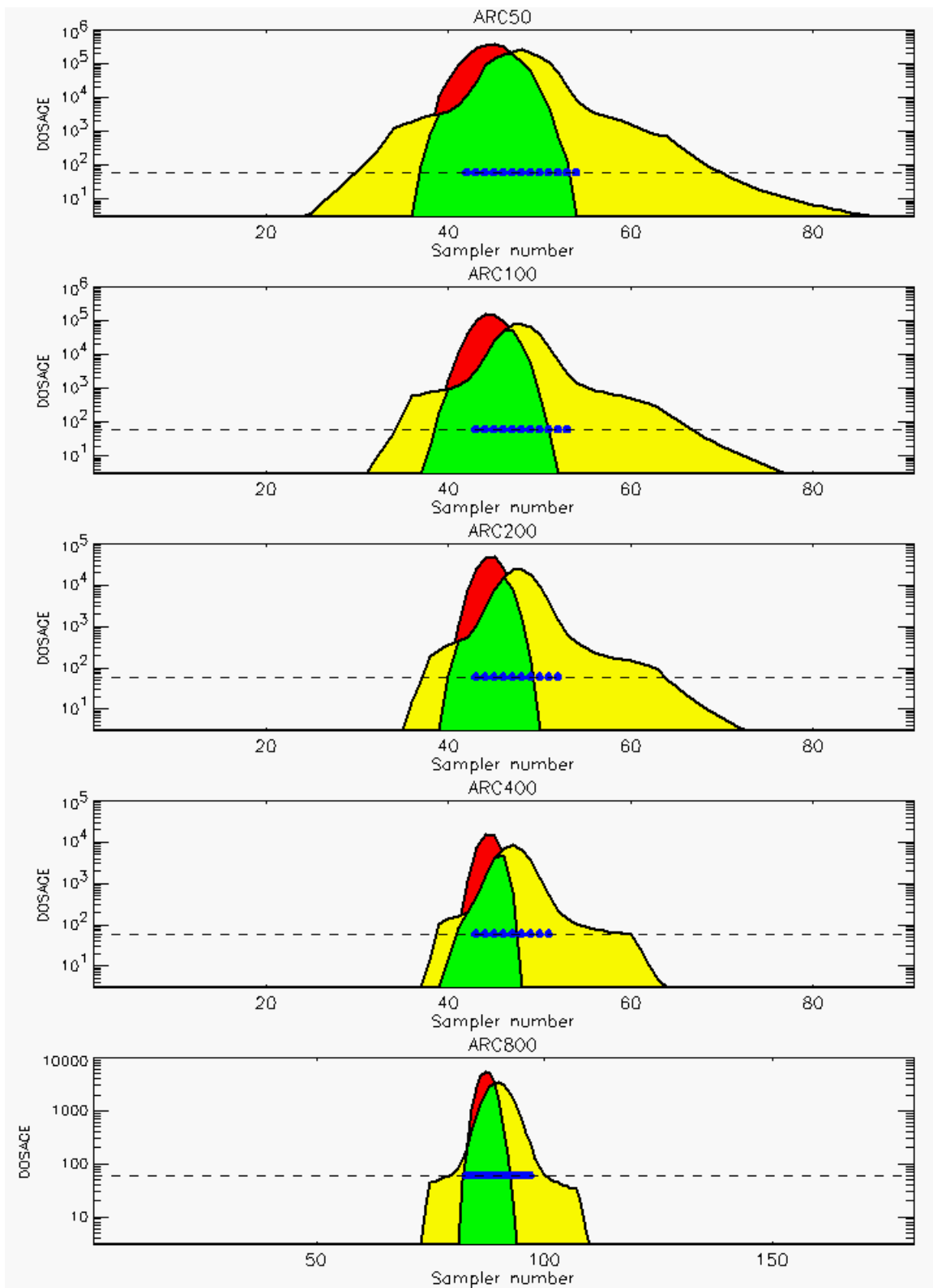


Figure D-12. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 17: Stability Class is 5

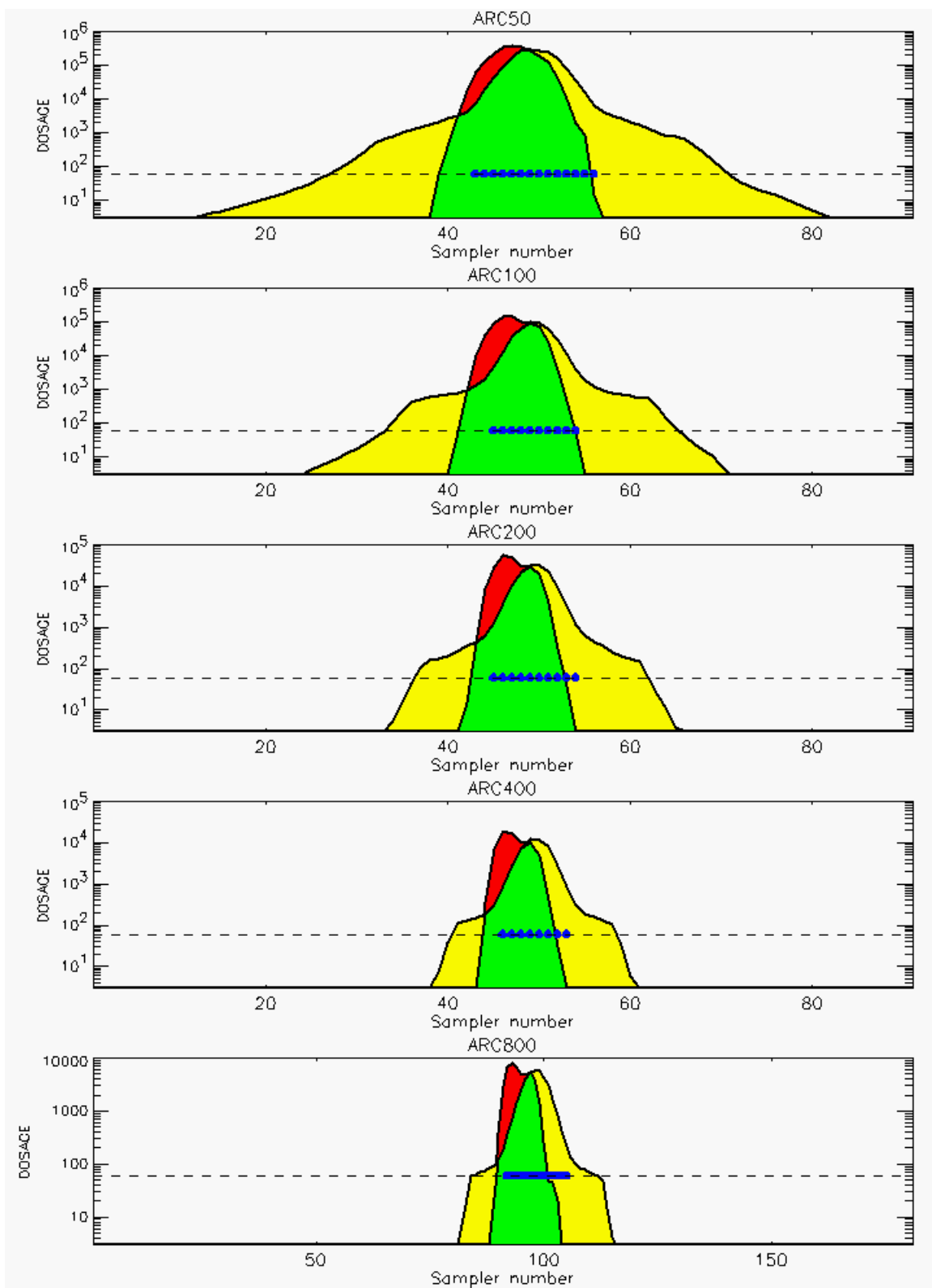


Figure D-13. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 18: Stability Class is 5

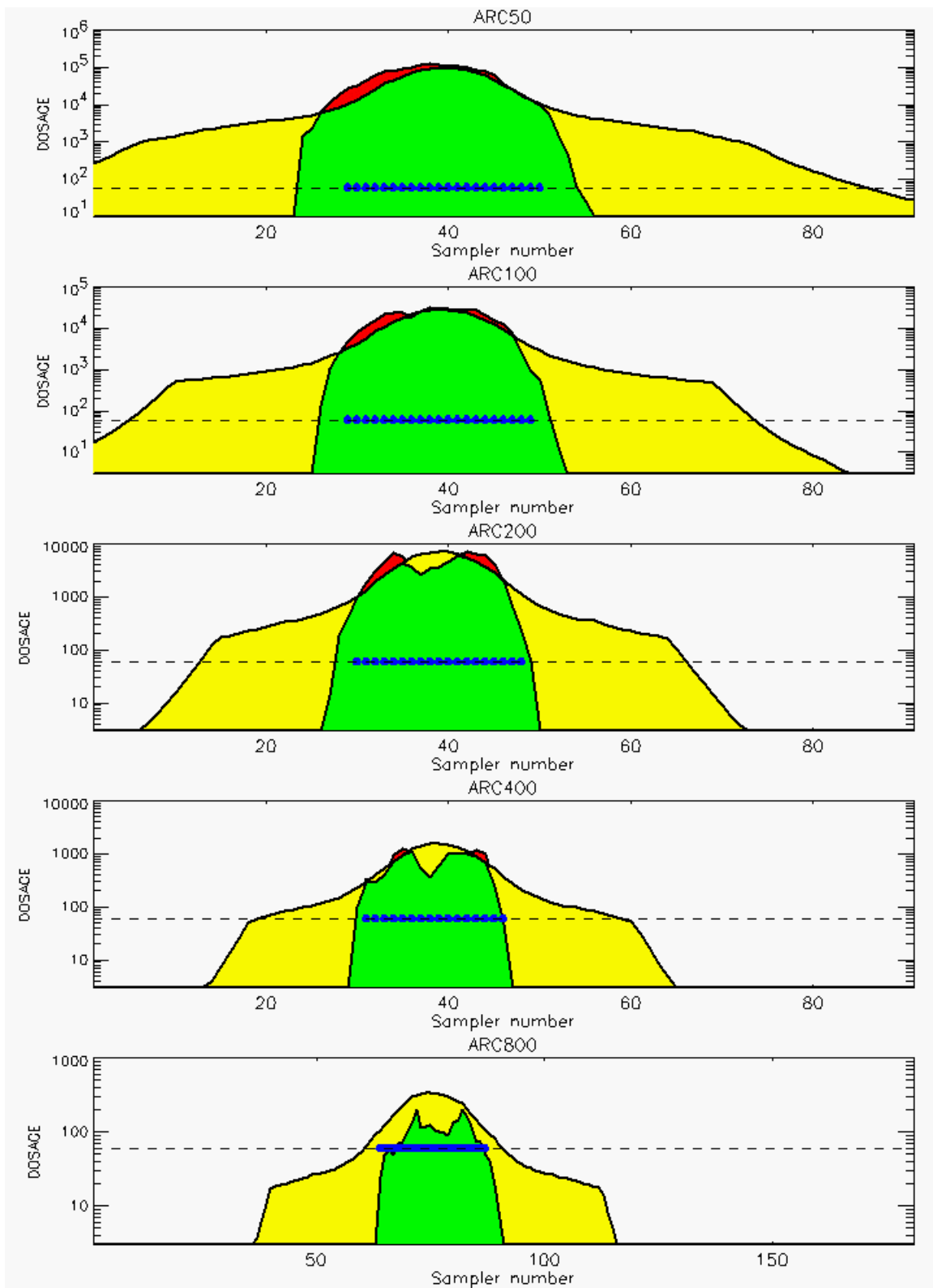


Figure D-14. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 19: Stability Class is 2

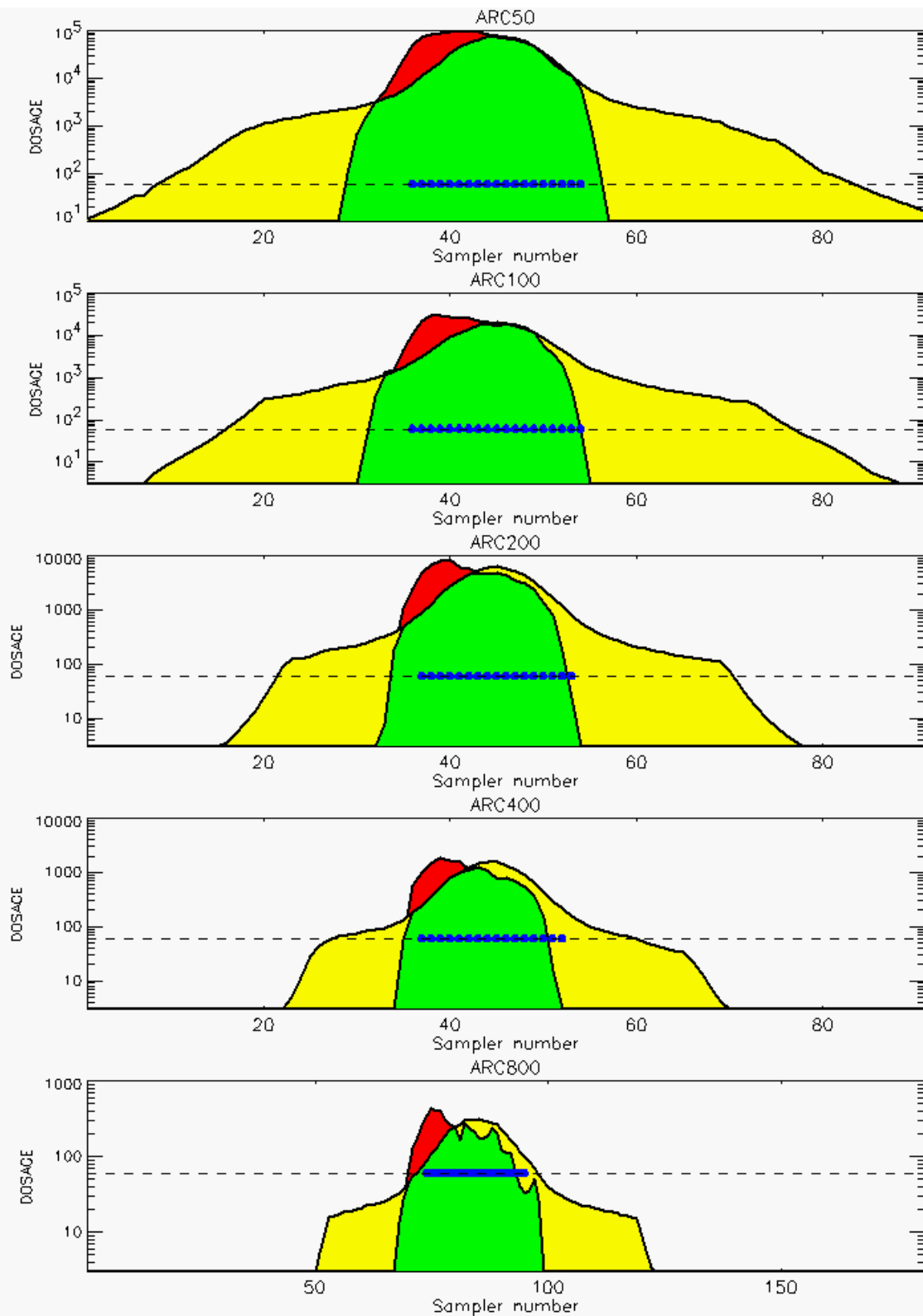


Figure D-15. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 20: Stability Class is 3

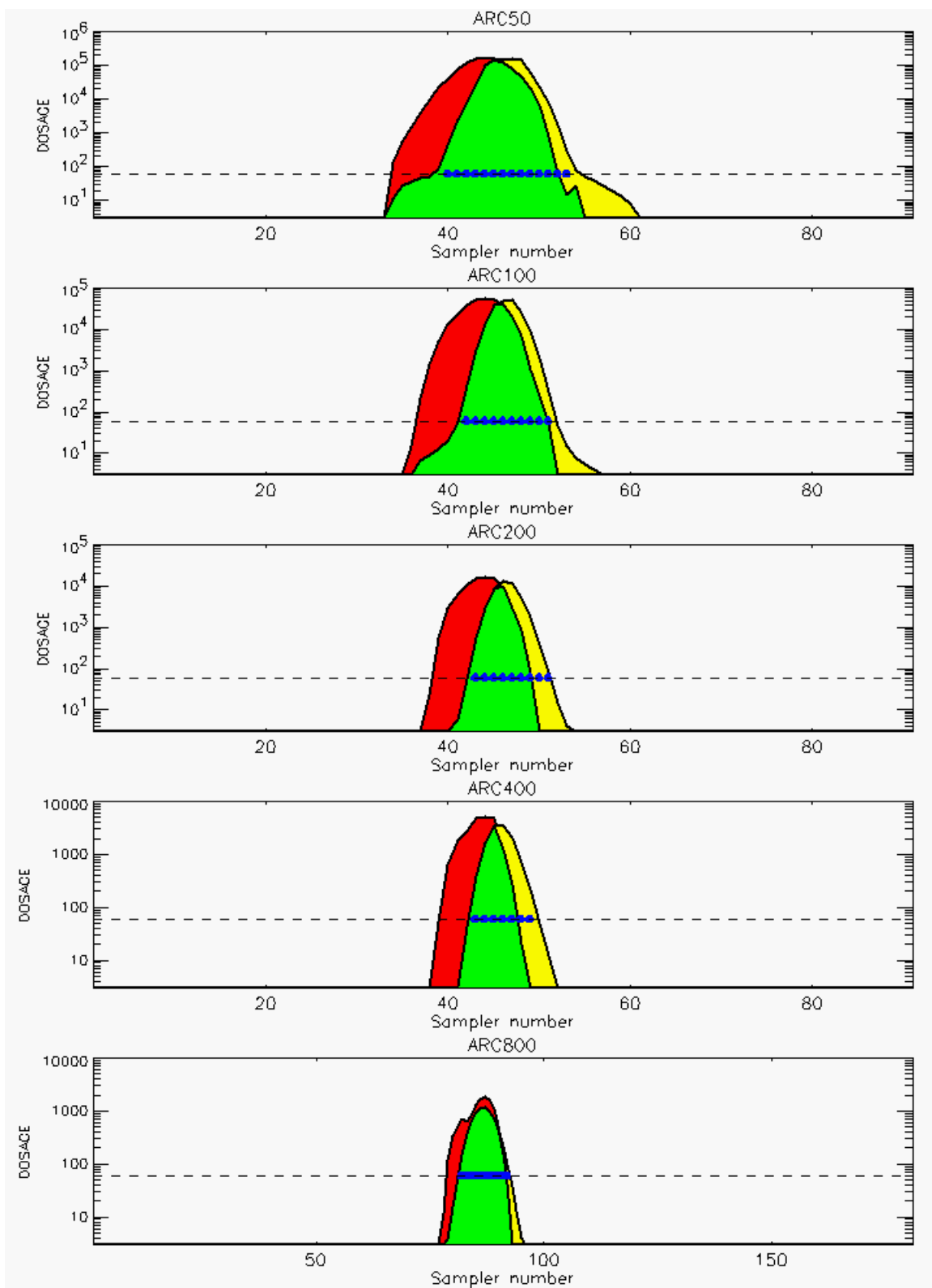


Figure D-16. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 21: Stability Class is 4

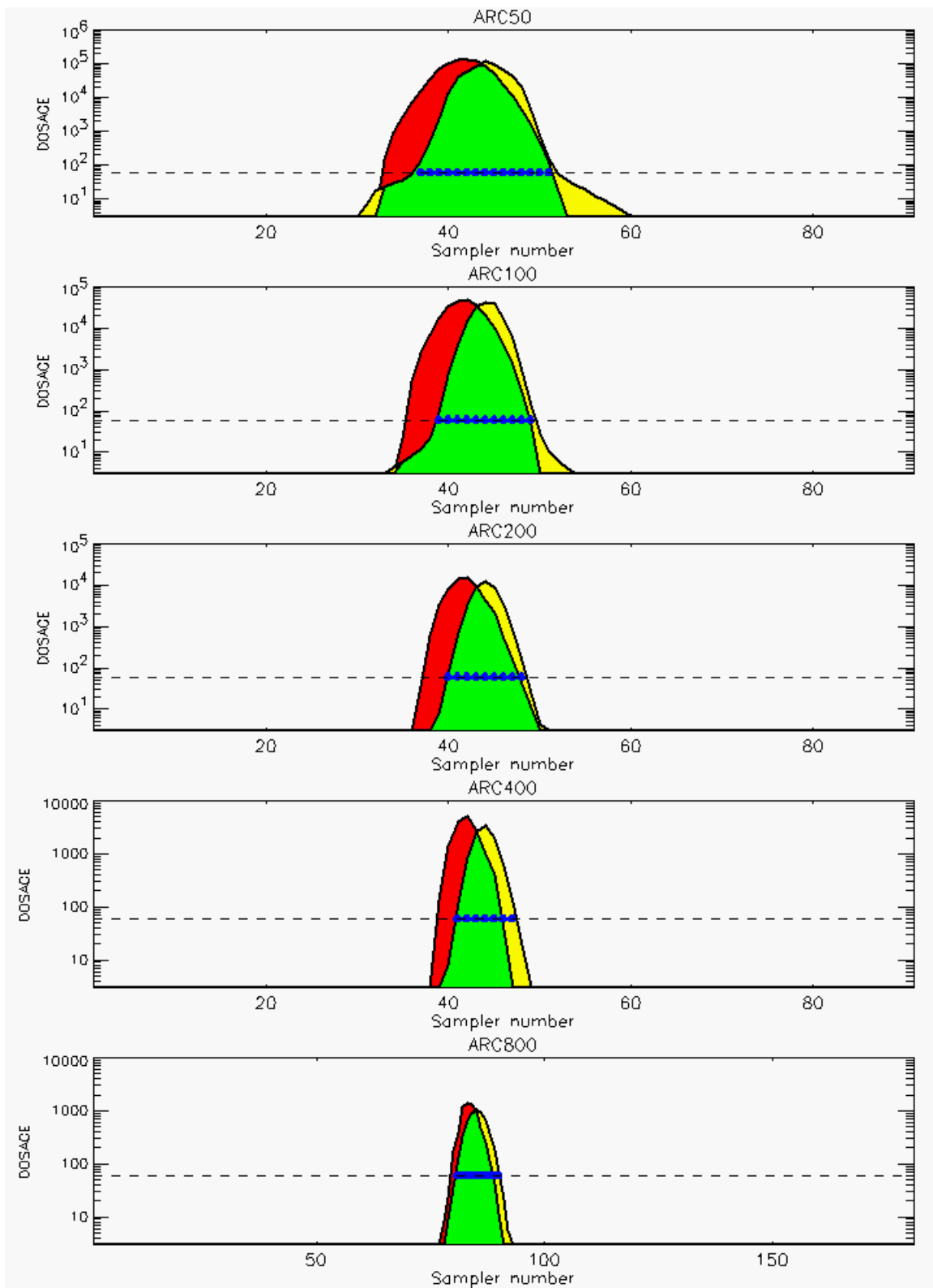


Figure D-17. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 22: Stability Class is 4

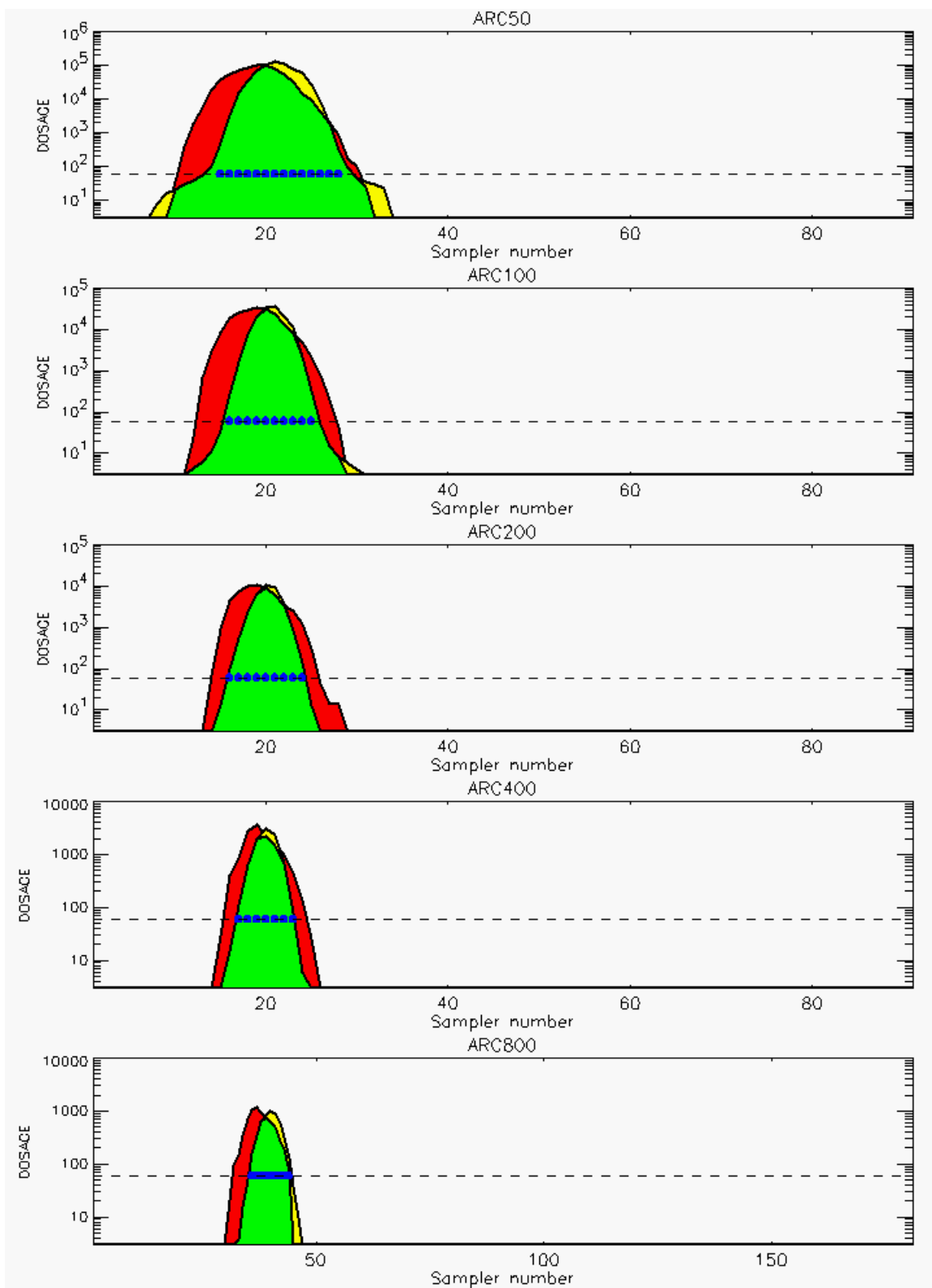


Figure D-18. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 23: Stability Class is 4

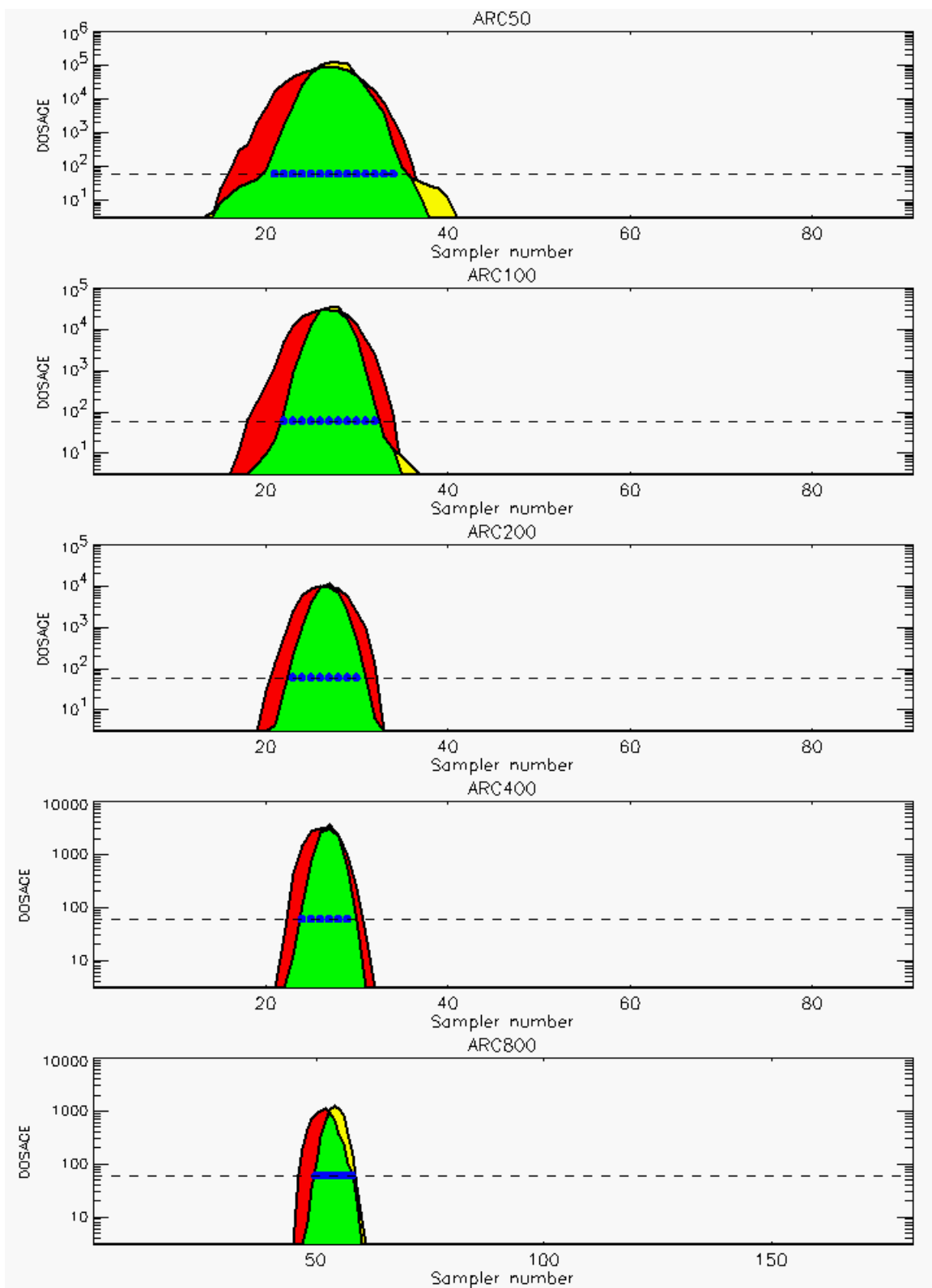


Figure D-19. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 24: Stability Class is 4

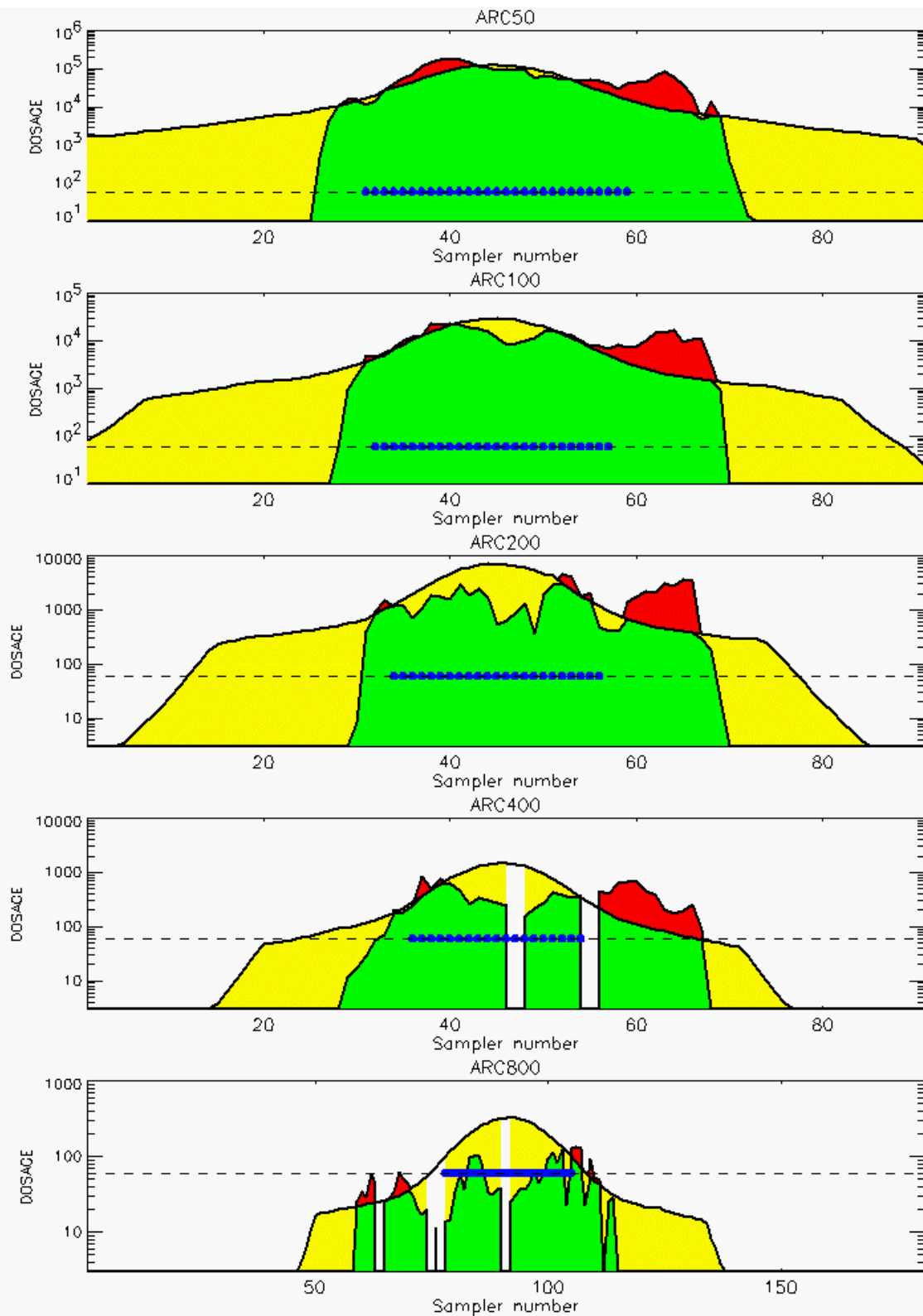


Figure D-20. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 25: Stability Class is 1

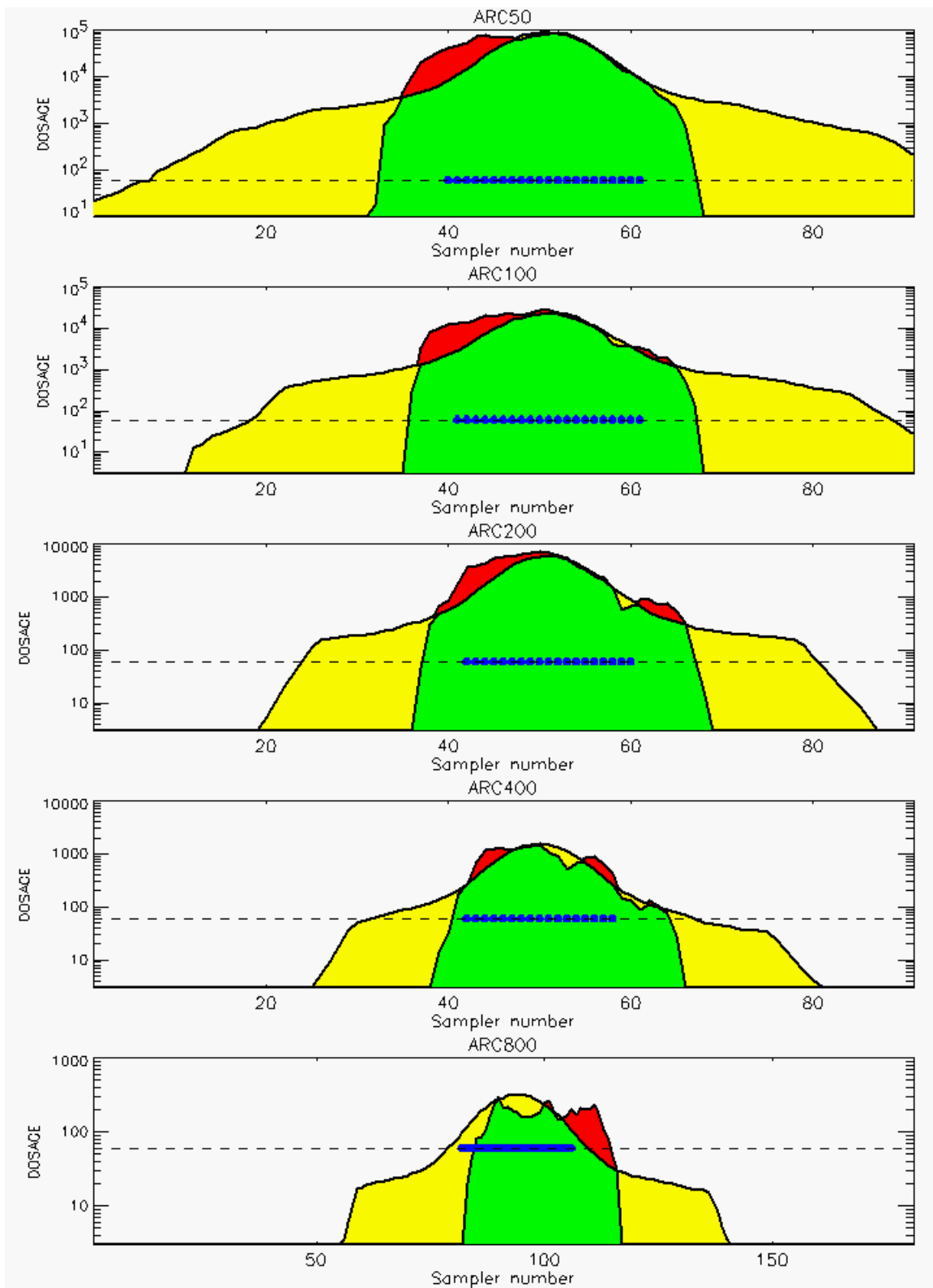


Figure D-21. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 26: Stability Class is 2

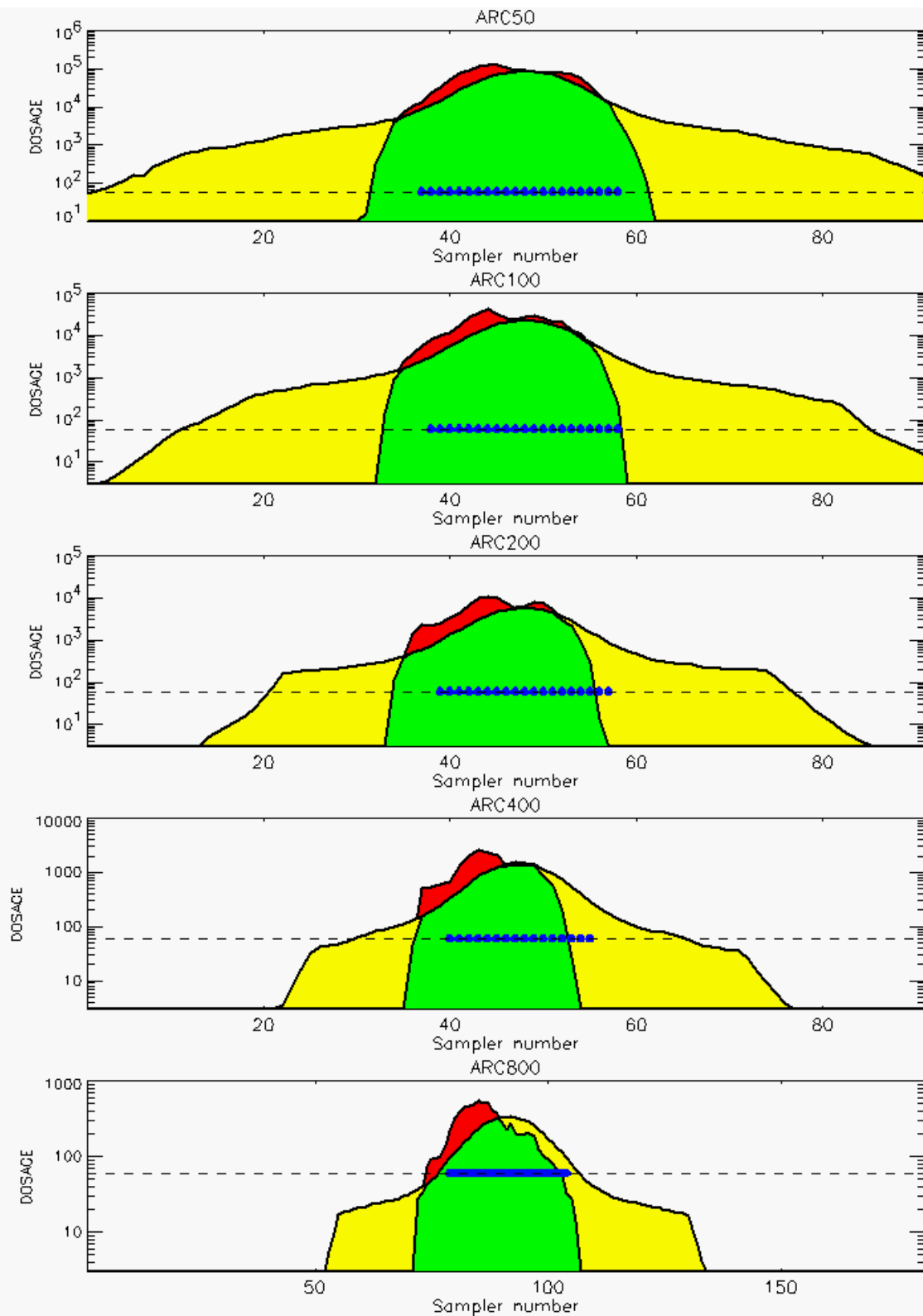


Figure D-22. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 27: Stability Class is 2

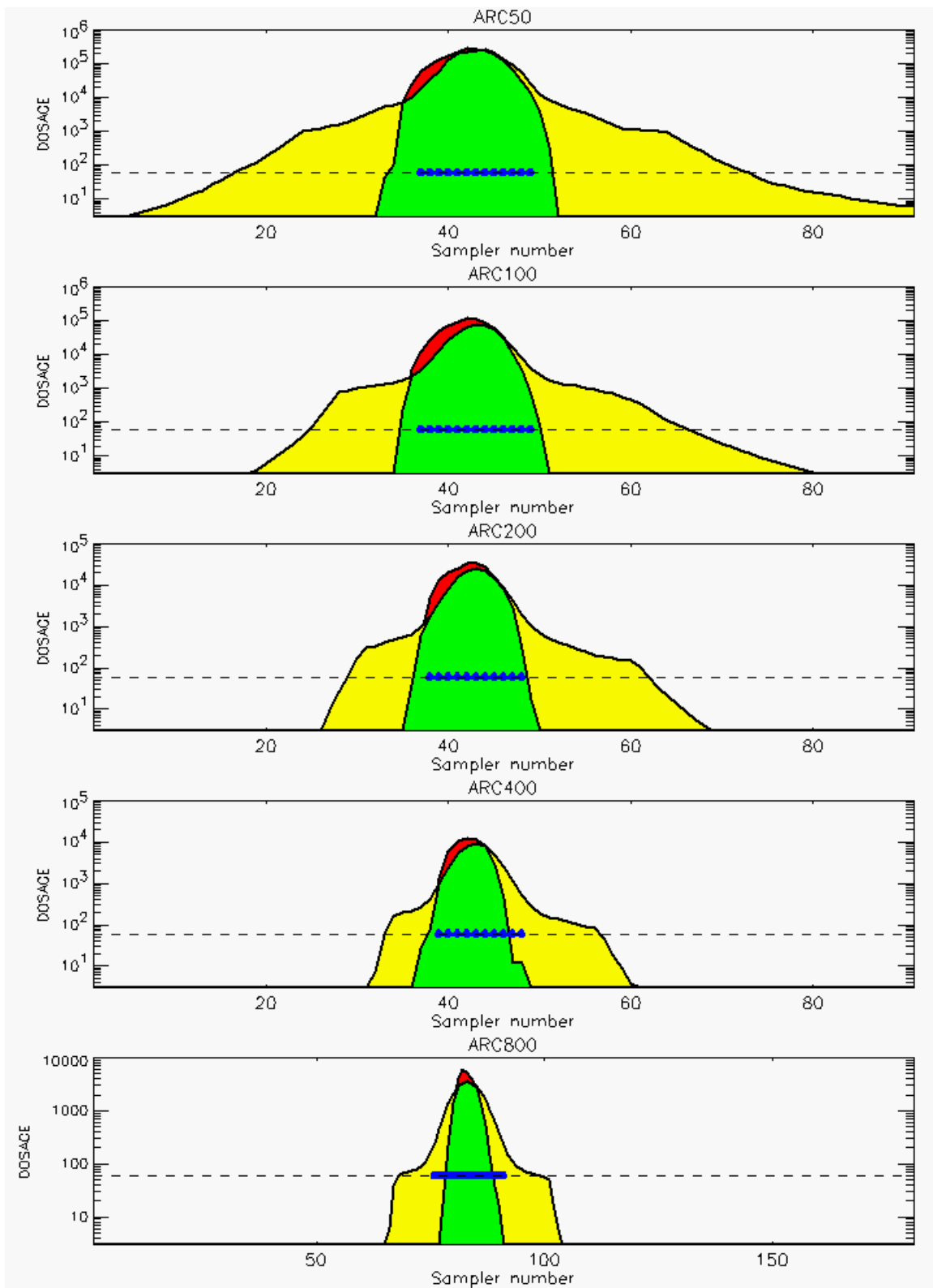


Figure D-23. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 28: Stability Class is 5

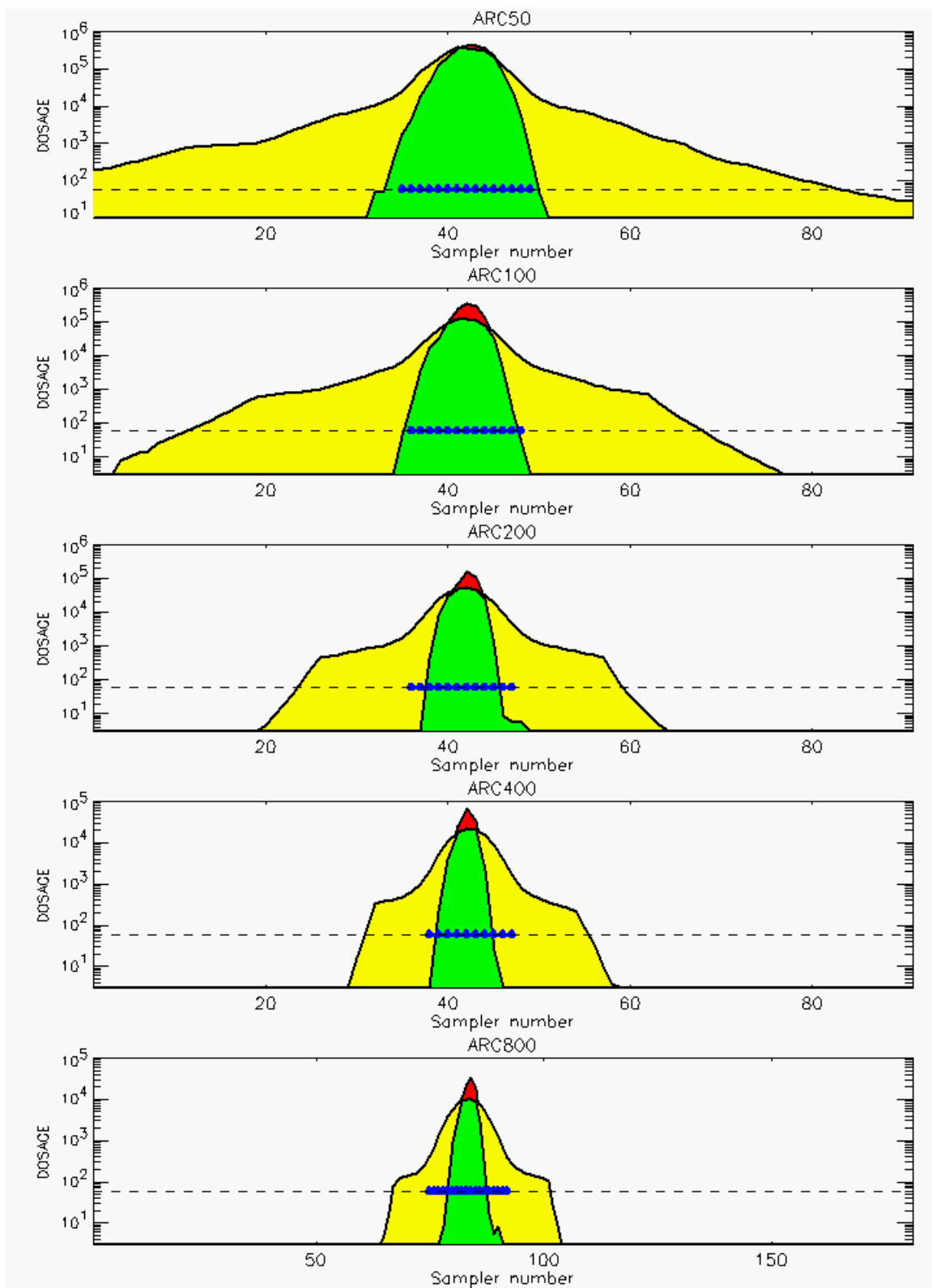


Figure D-24. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 32: Stability Class is 6

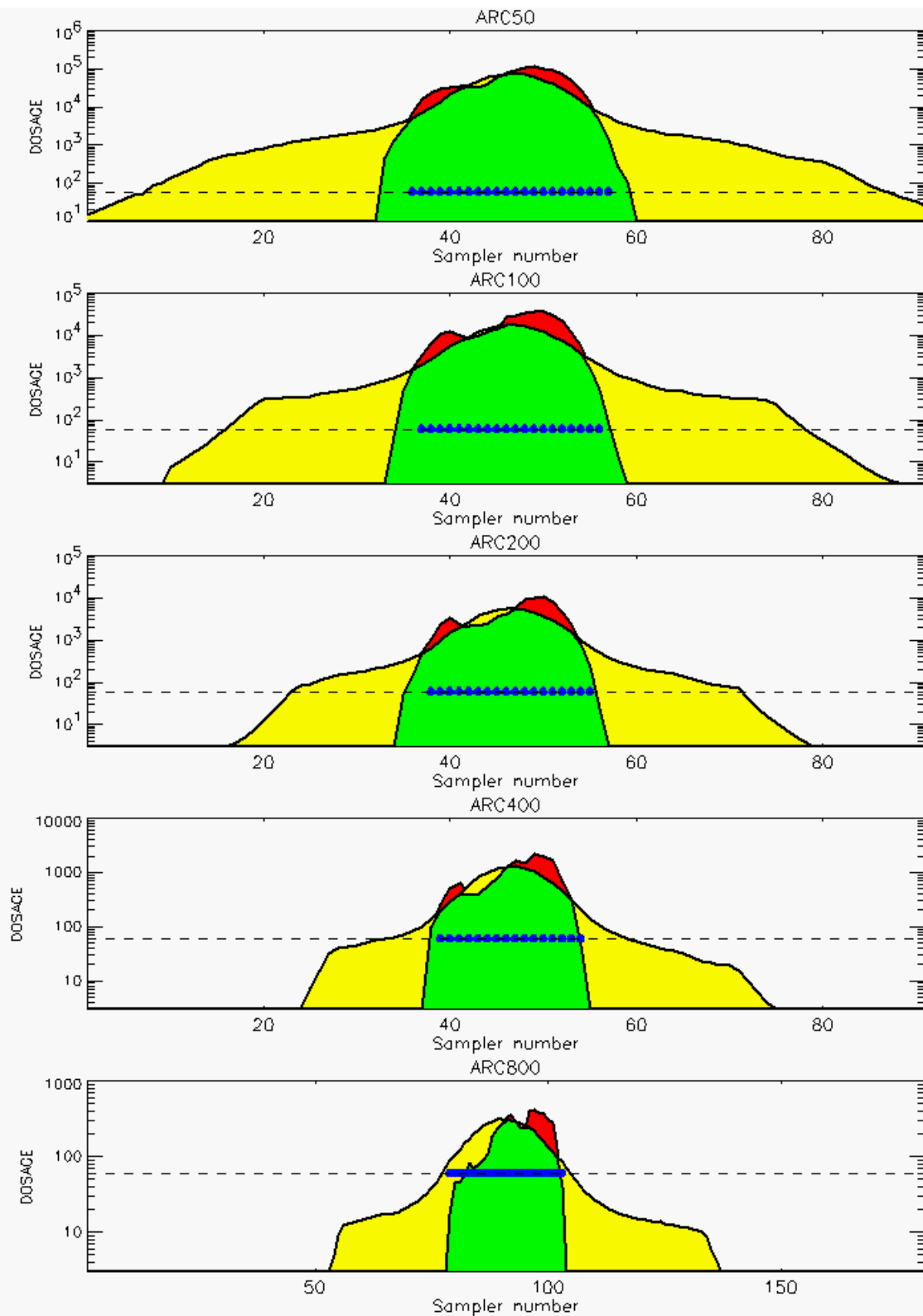


Figure D-25. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 33: Stability Class is 3

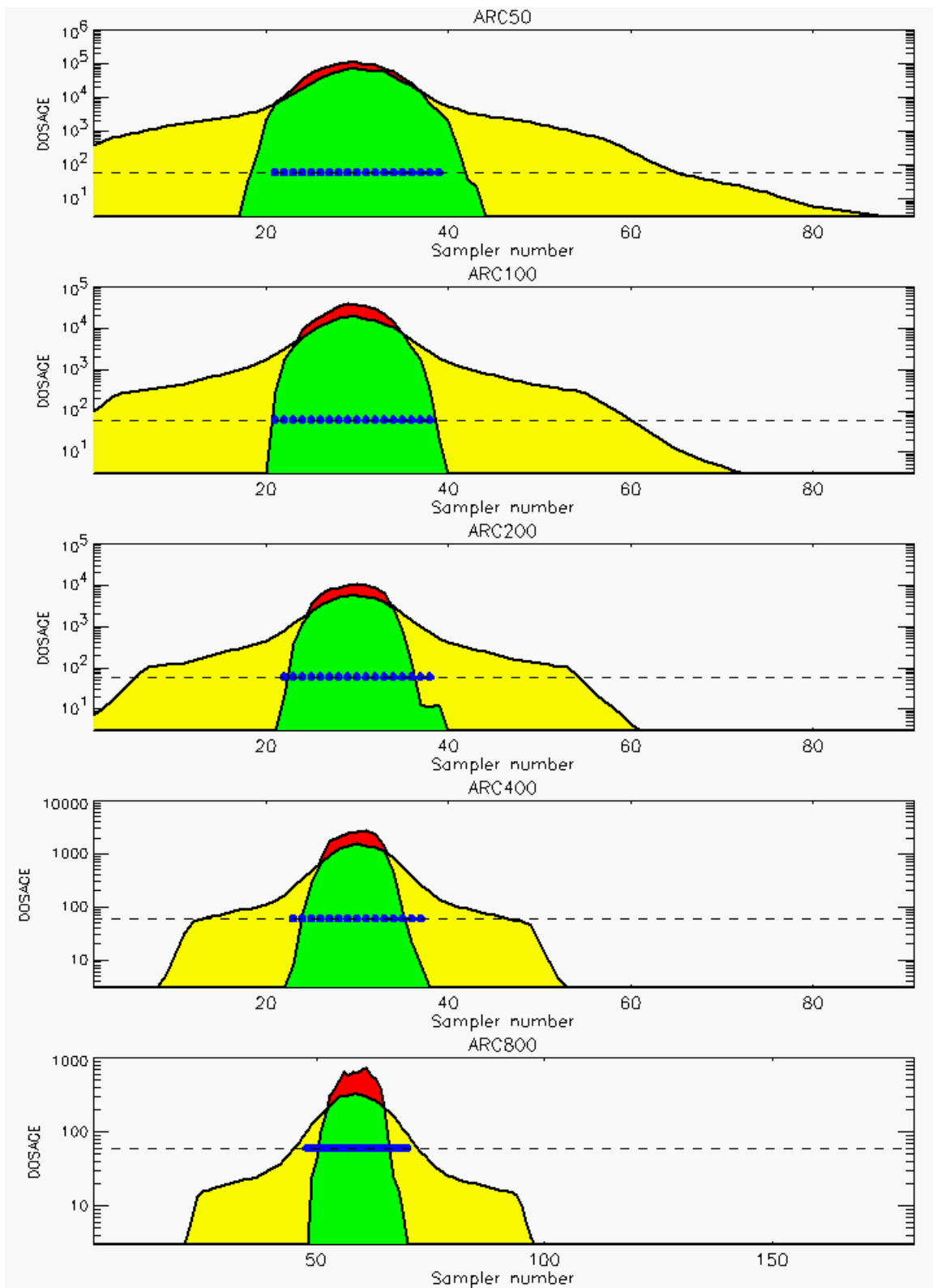


Figure D-26. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 34: Stability Class is 3

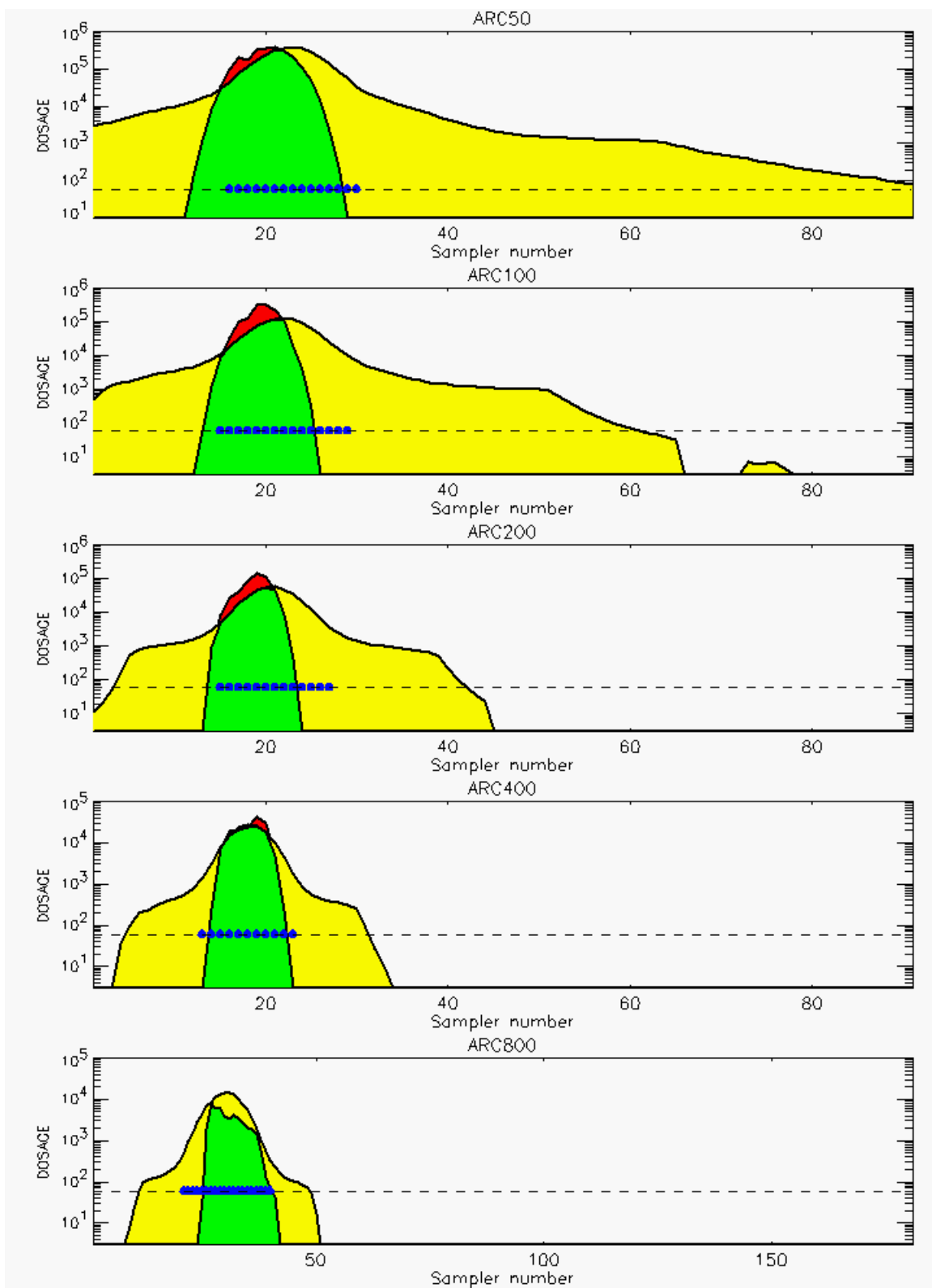


Figure D-27. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 35: Stability Class is 6

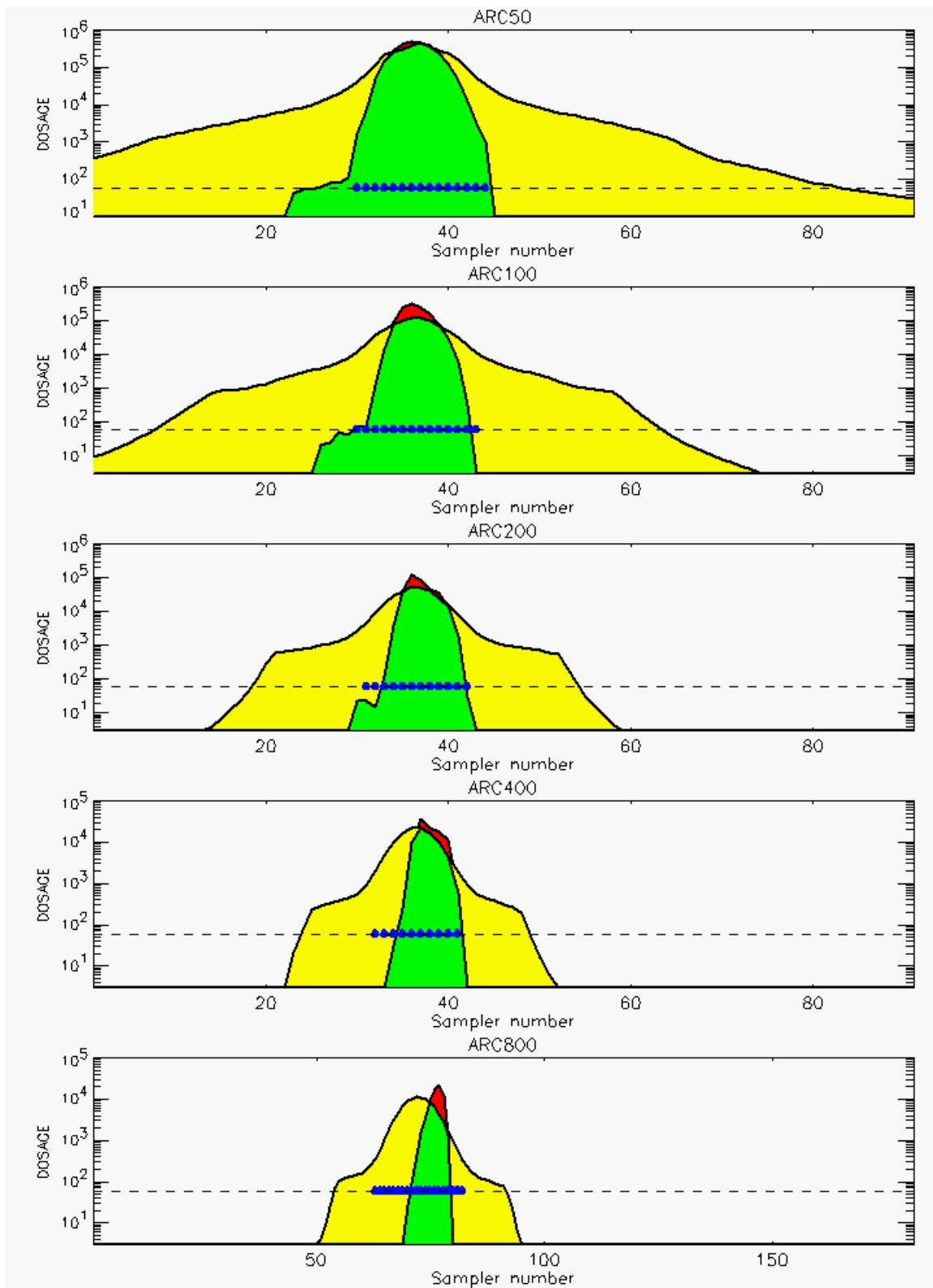


Figure D-28. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 36: Stability Class is 6

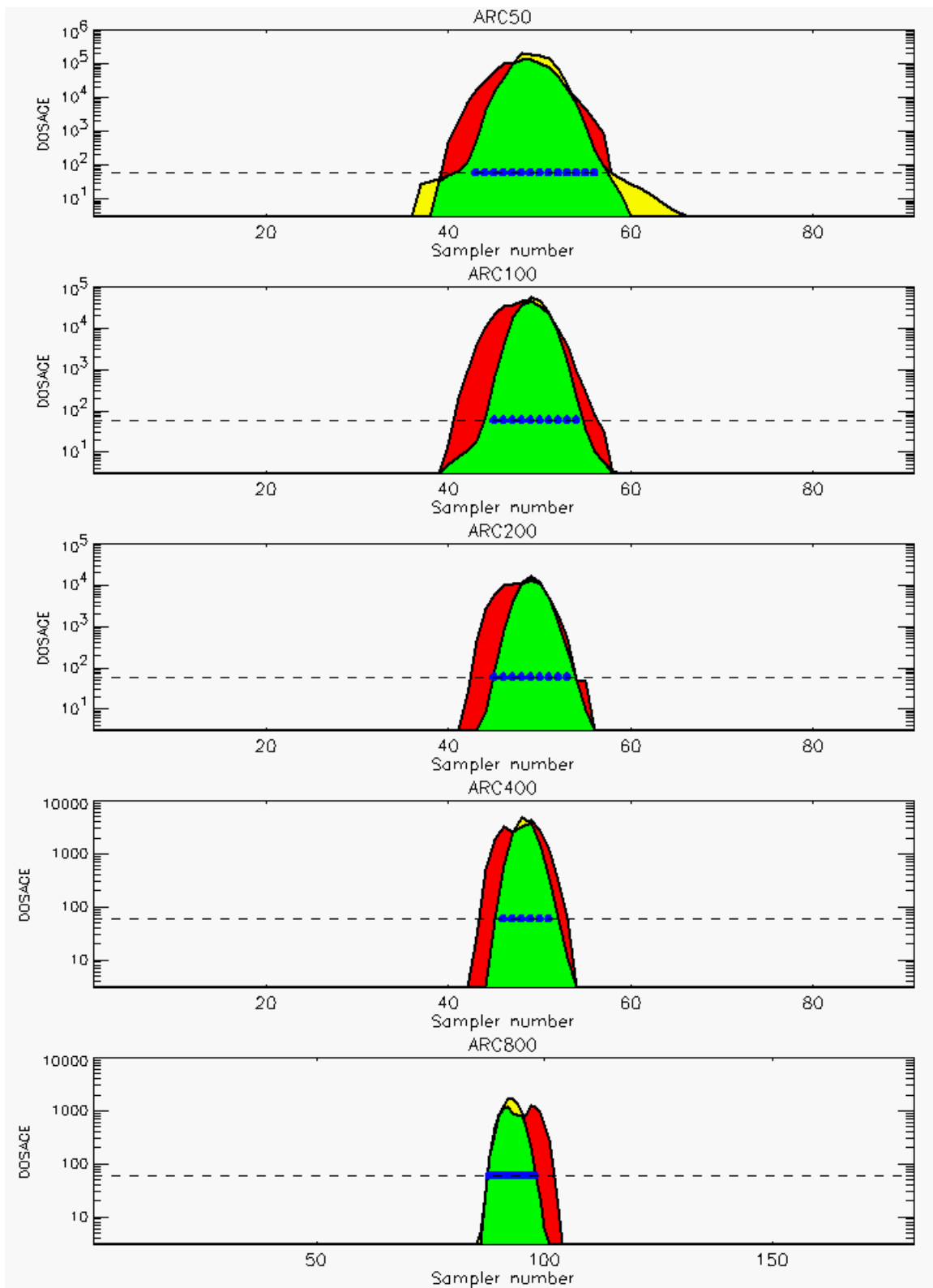


Figure D-29. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 37: Stability Class is 4

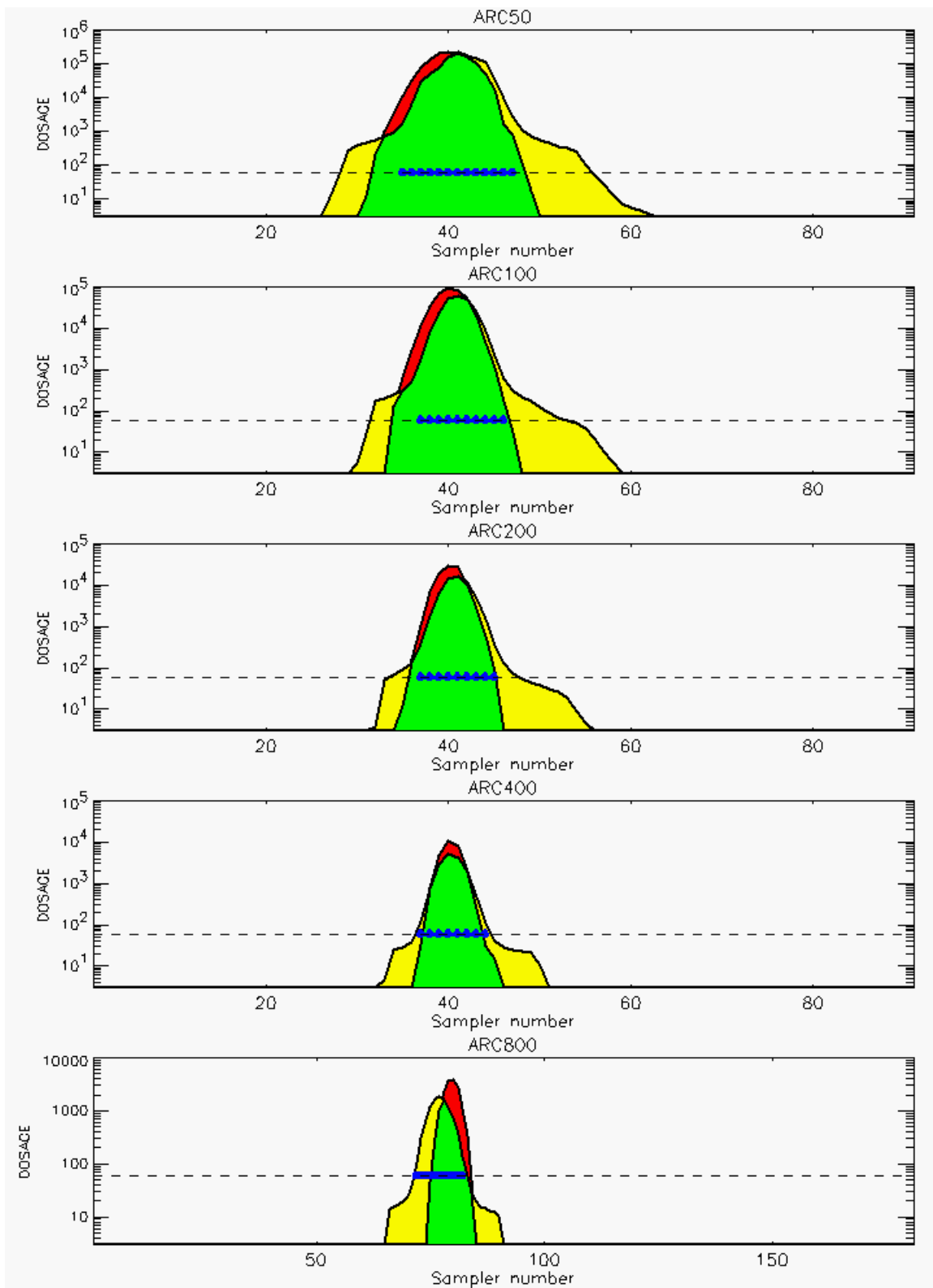


Figure D-30. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 38: Stability Class is 4

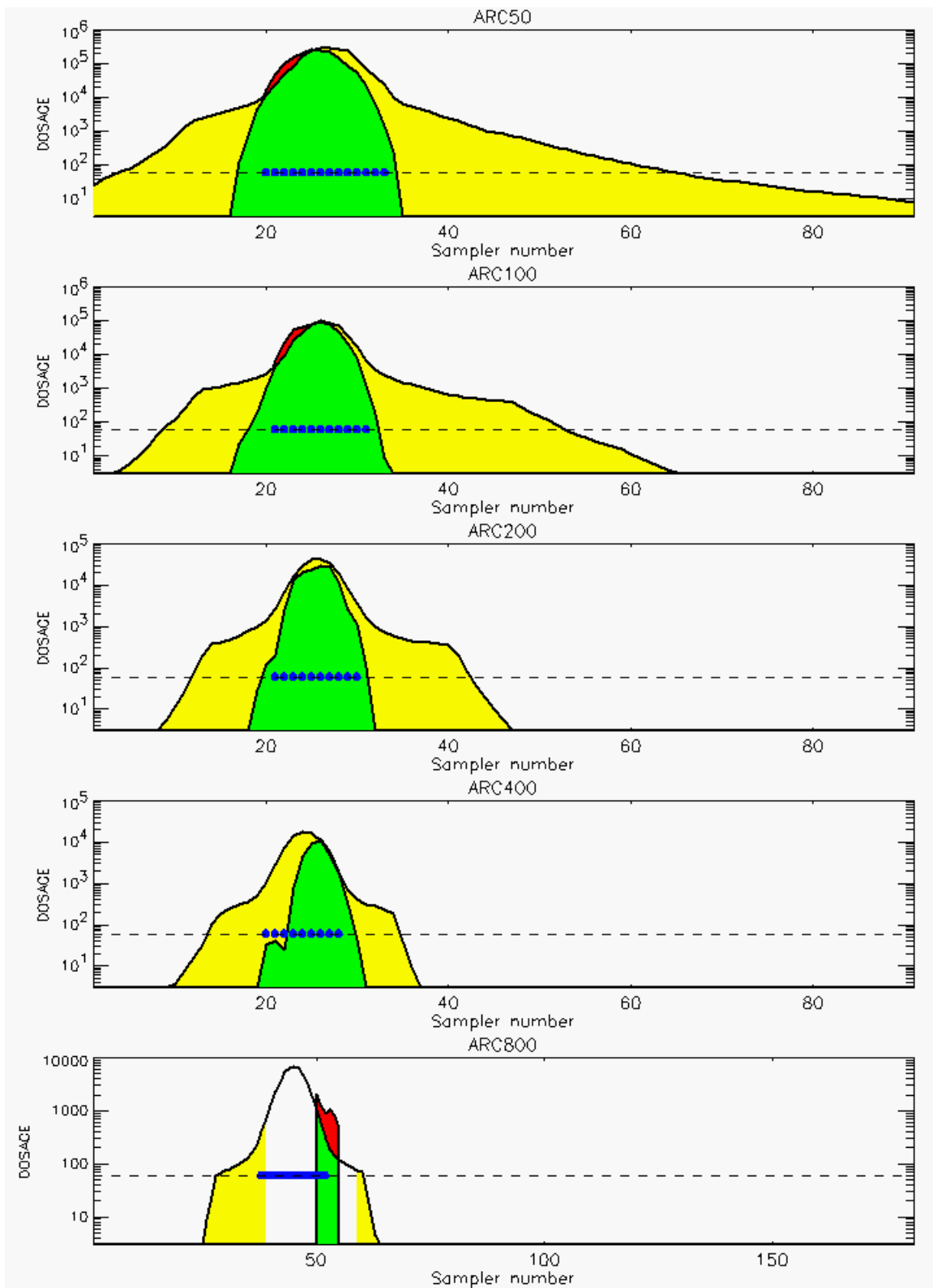


Figure D-31. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 39: Stability Class is 6

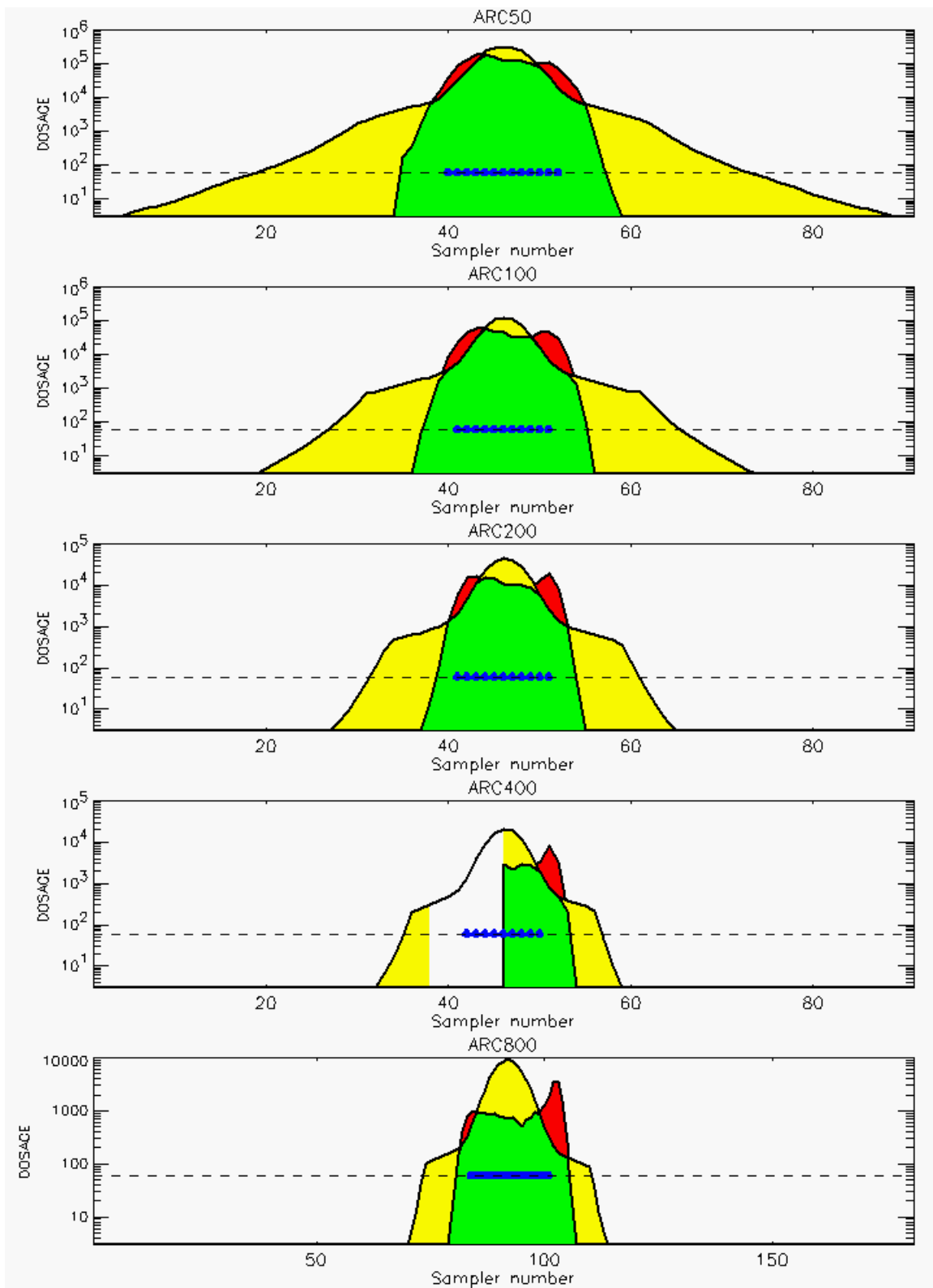


Figure D-32. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 40: Stability Class is 6

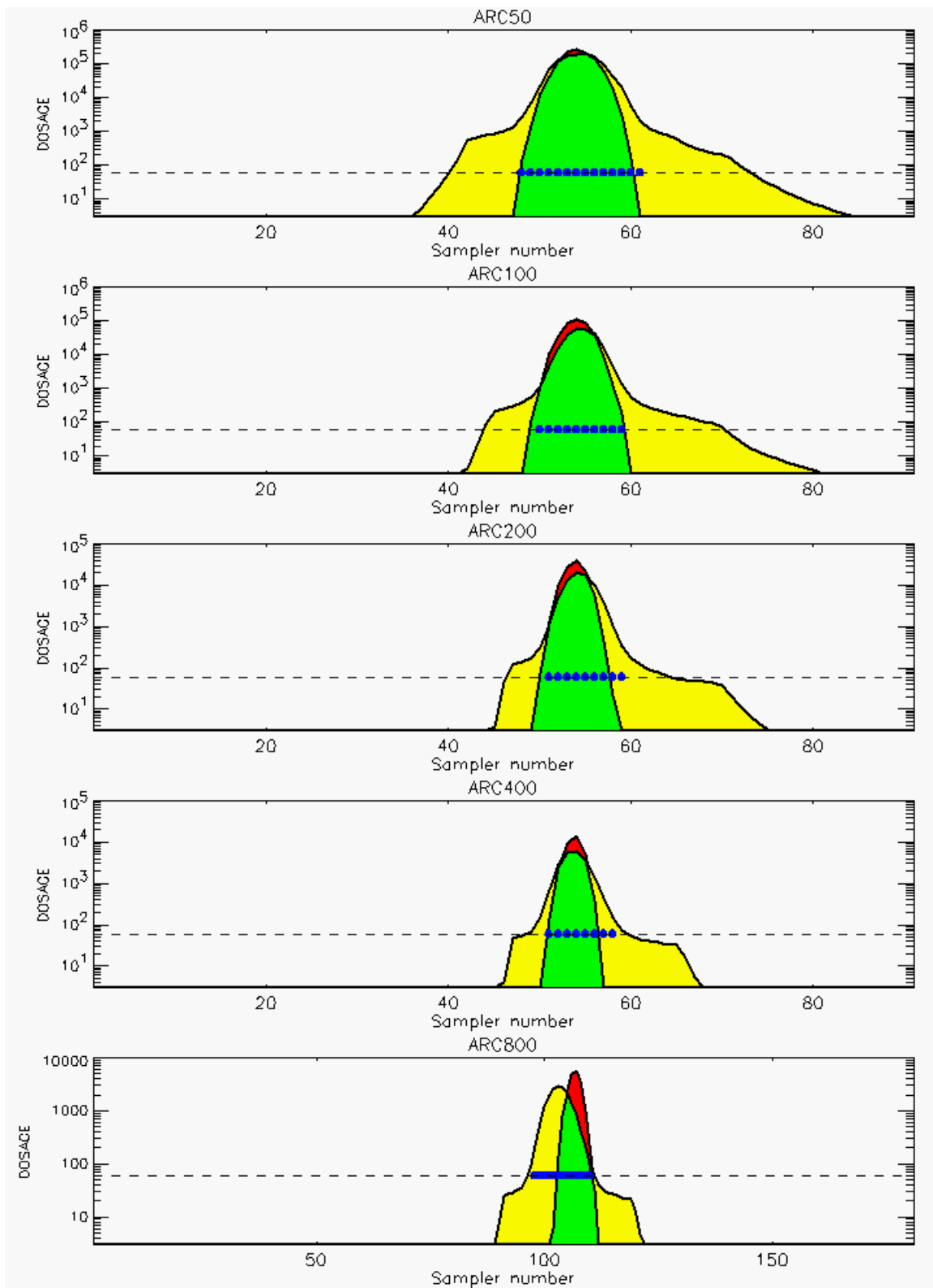


Figure D-33. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 41: Stability Class is 5

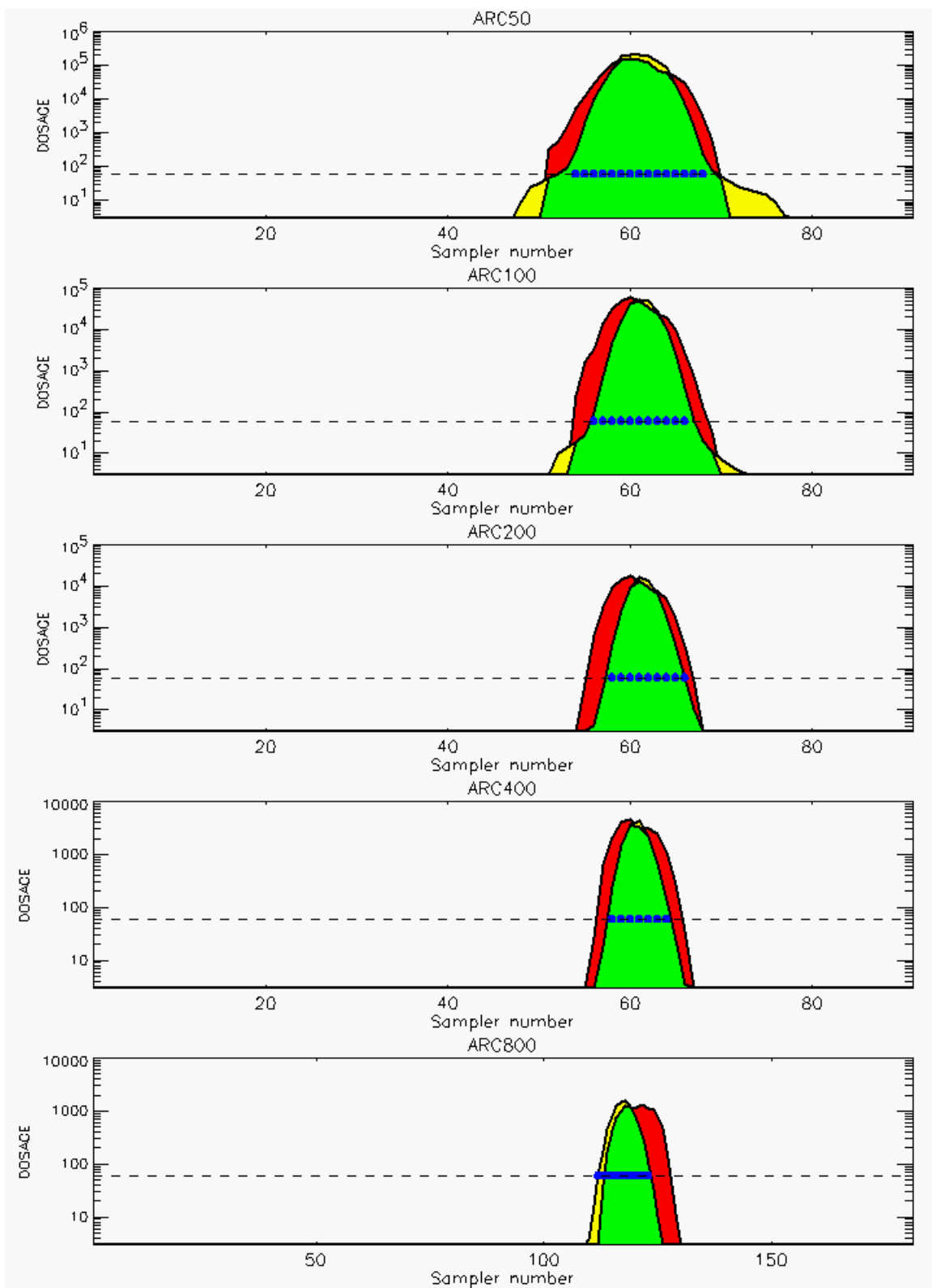


Figure D-34. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 42: Stability Class is 4

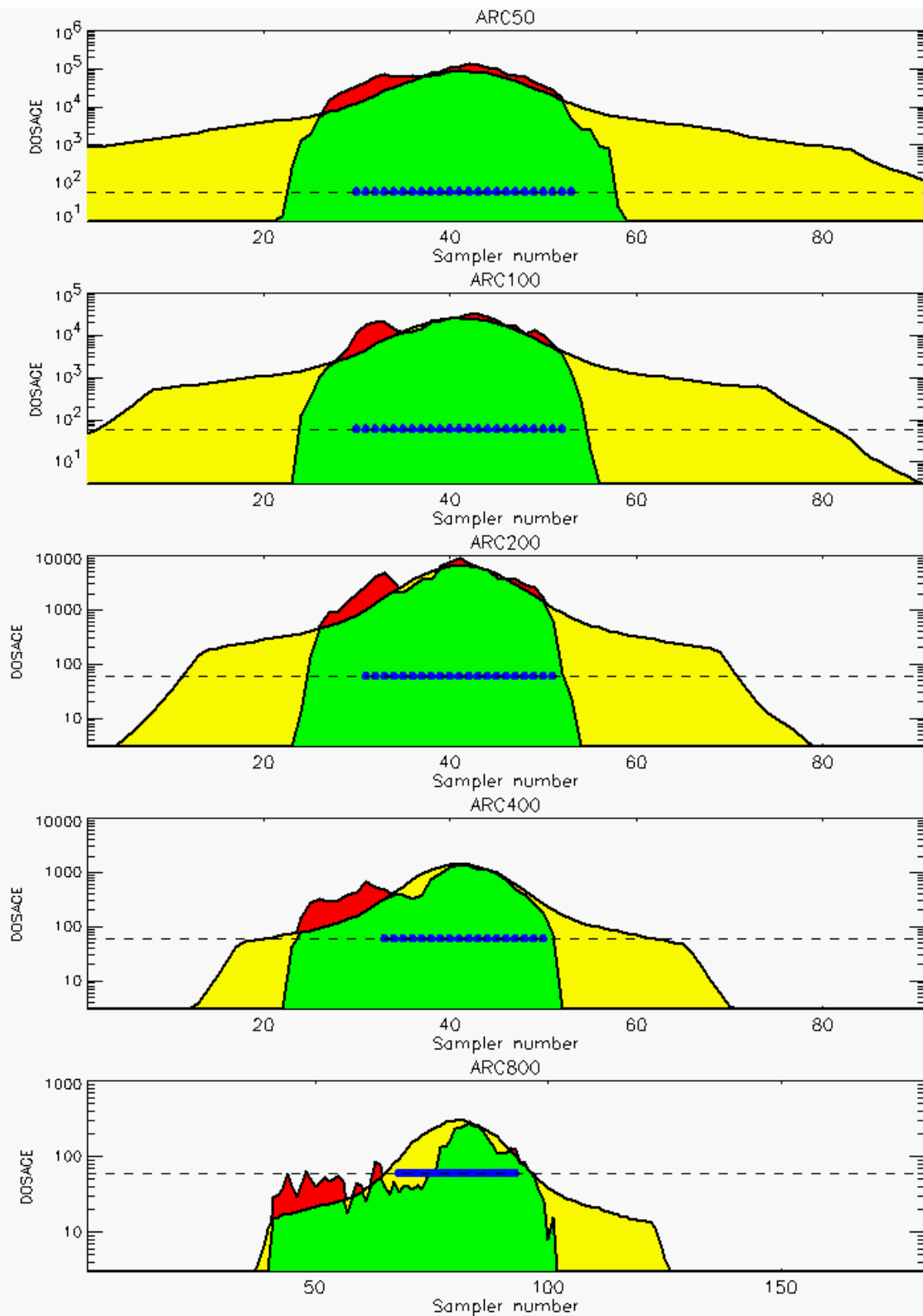


Figure D-35. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 43: Stability Class is 1

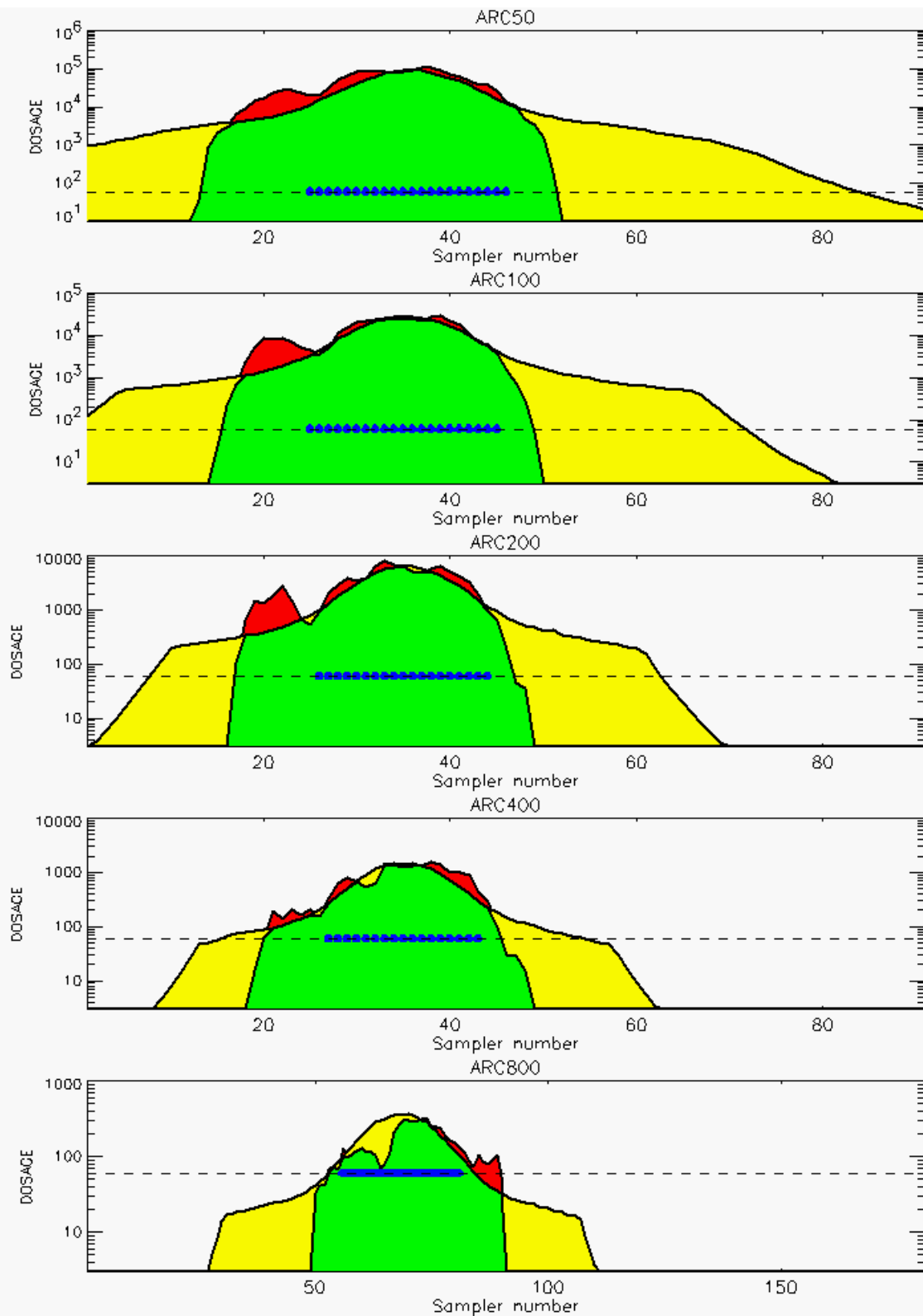


Figure D-36. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 44: Stability Class is 2

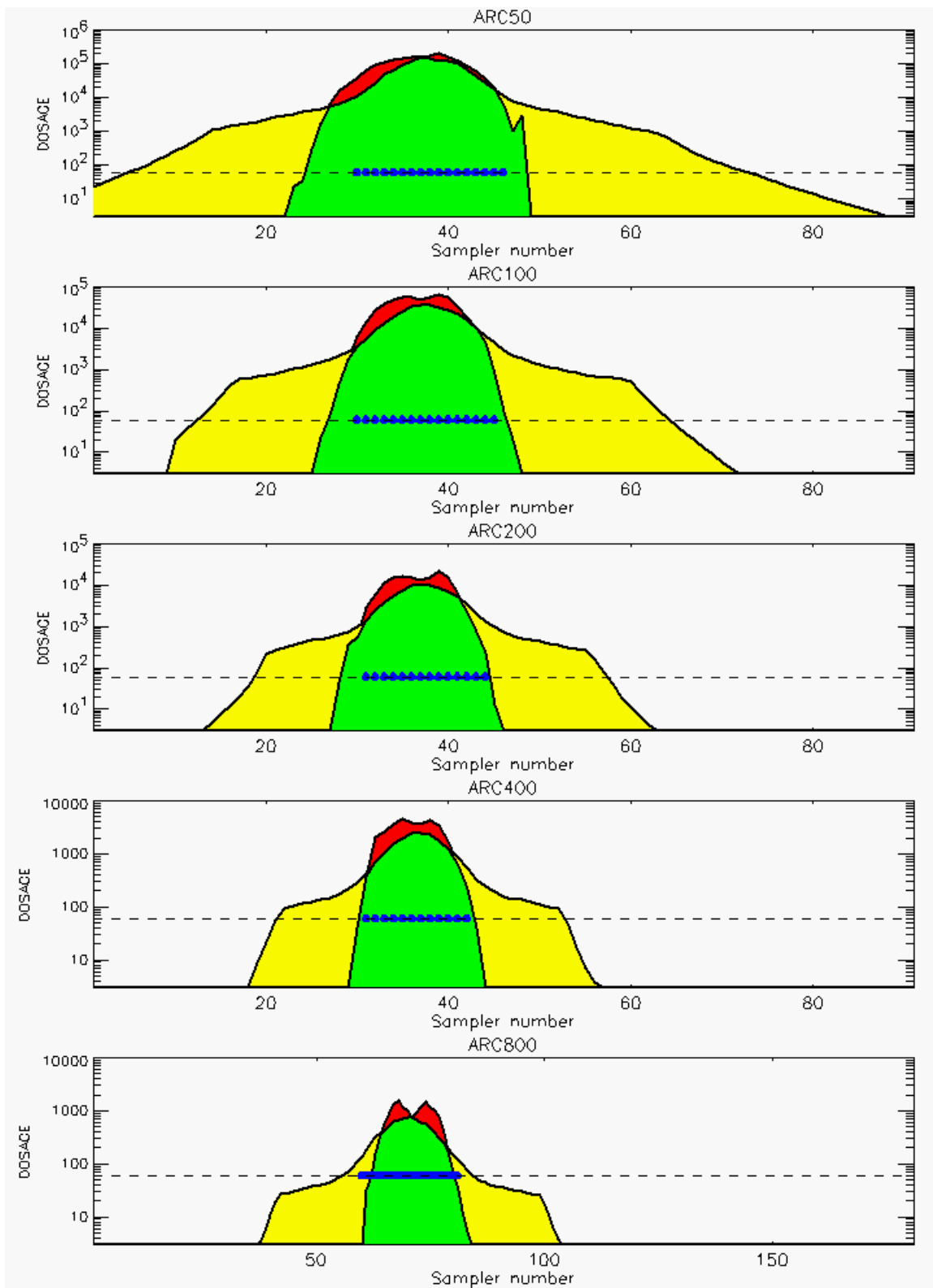


Figure D-37. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 45: Stability Class is 3

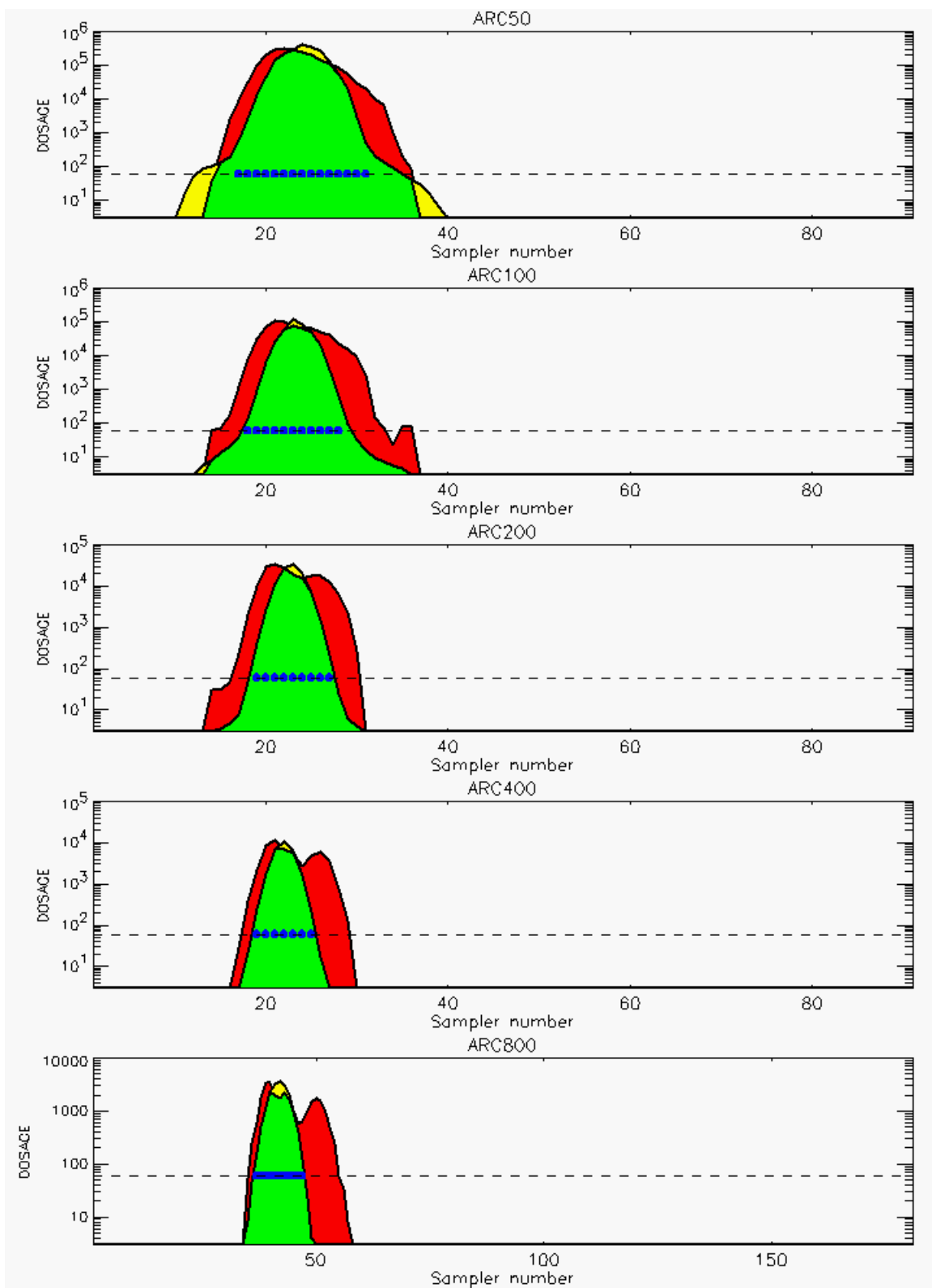


Figure D-38. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 46: Stability Class is 4

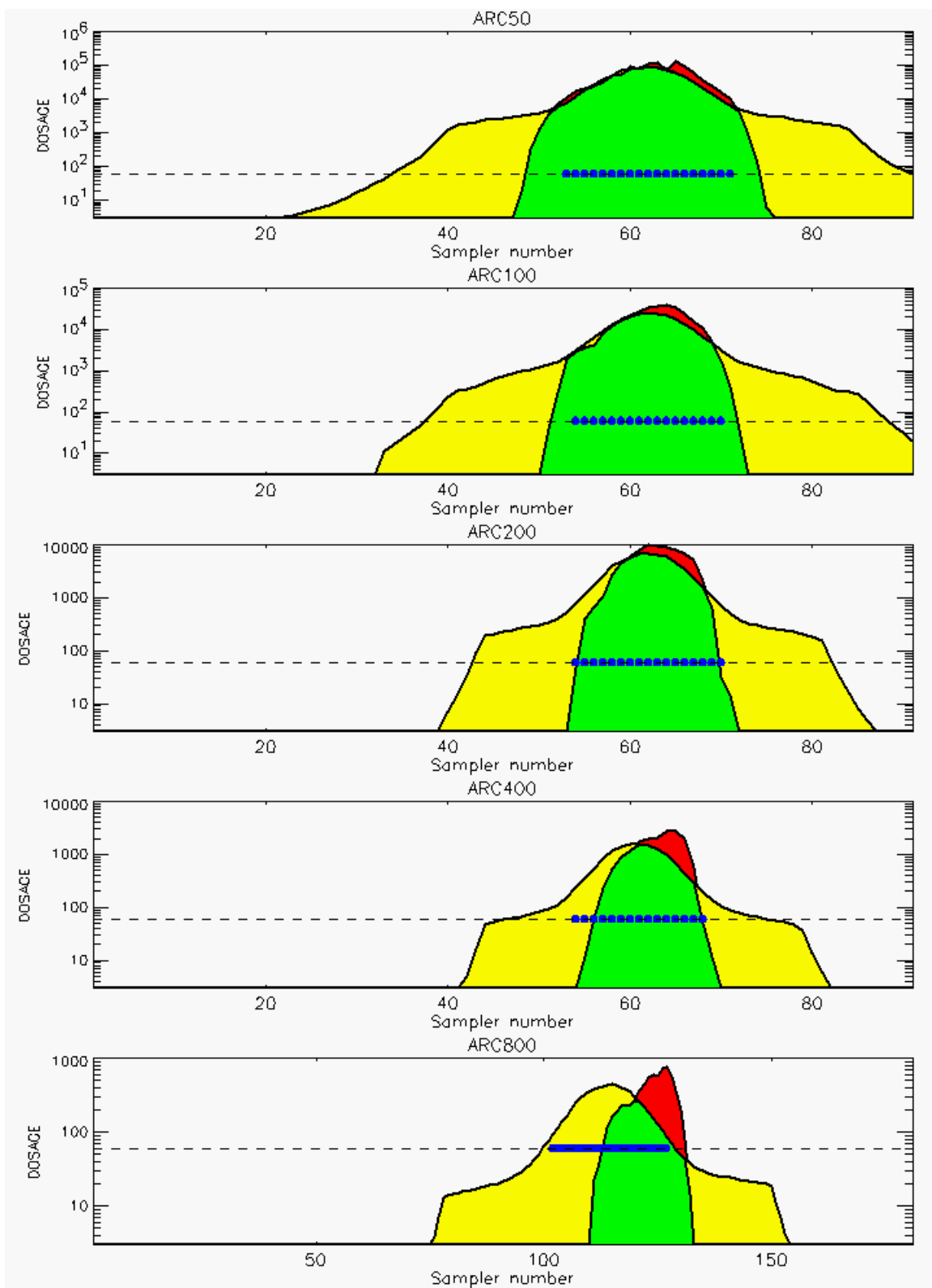


Figure D-39. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 48: Stability Class is 3

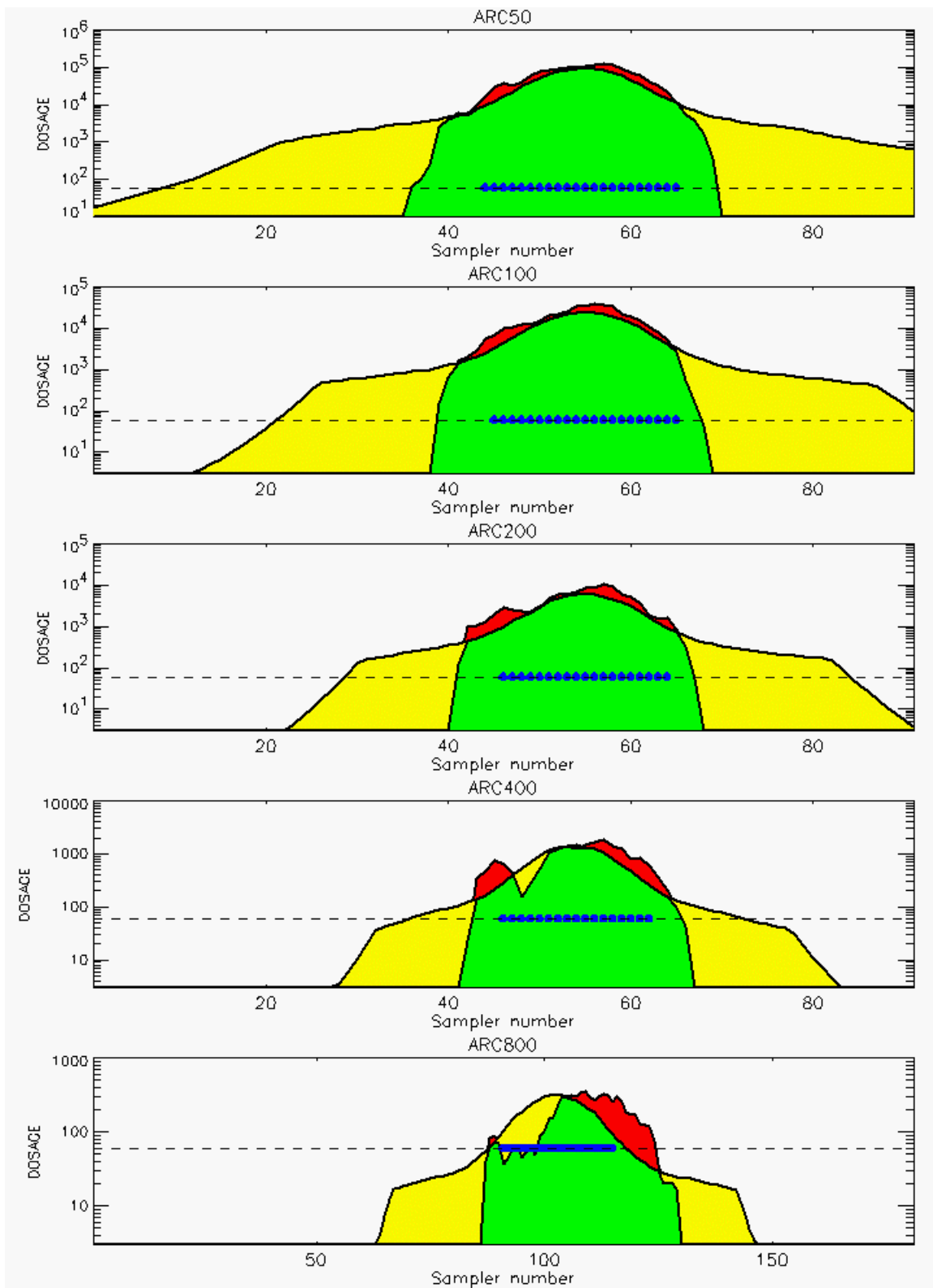


Figure D-40. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 49: Stability Class is 2

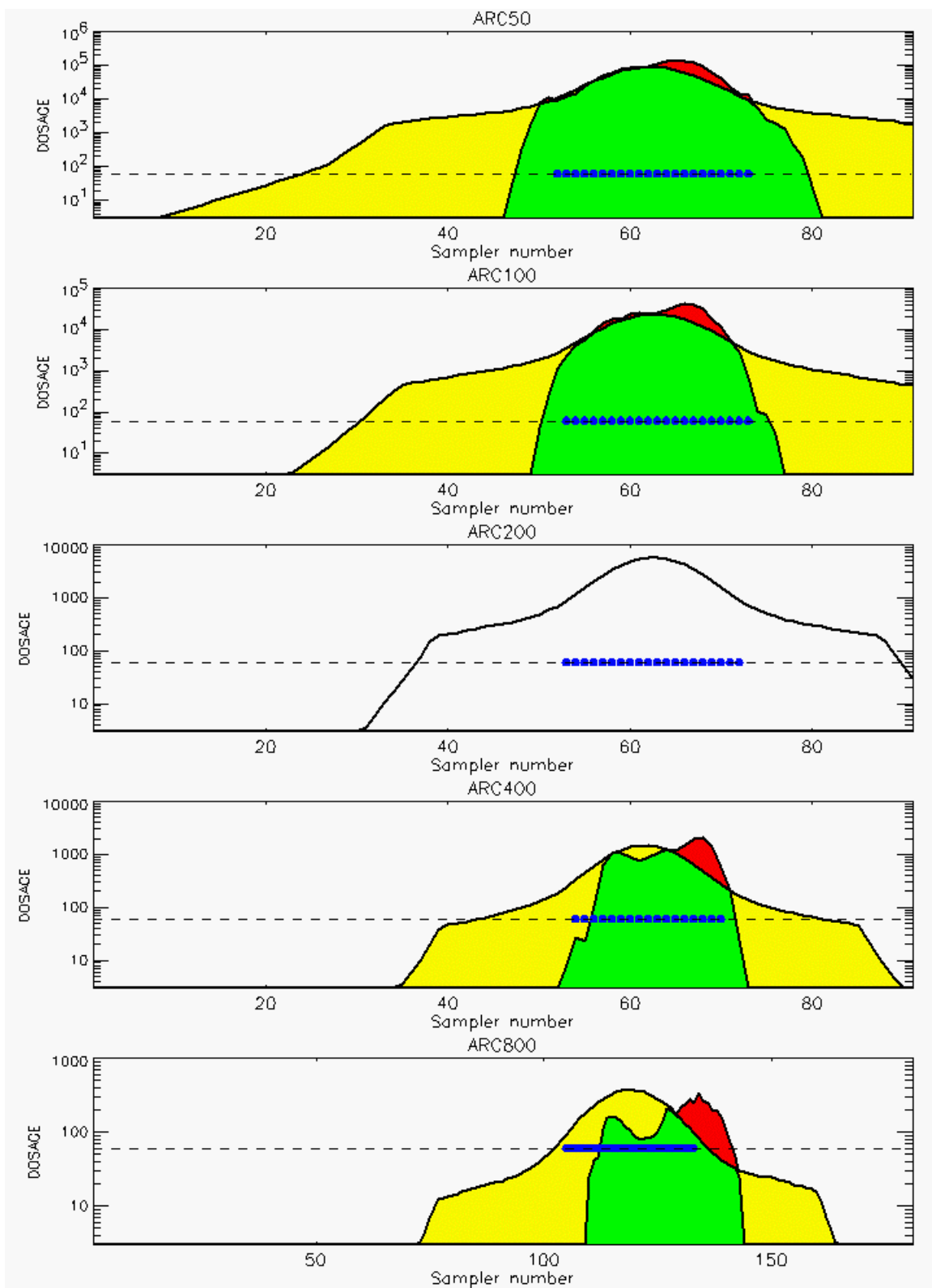


Figure D-41. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 50: Stability Class is 2

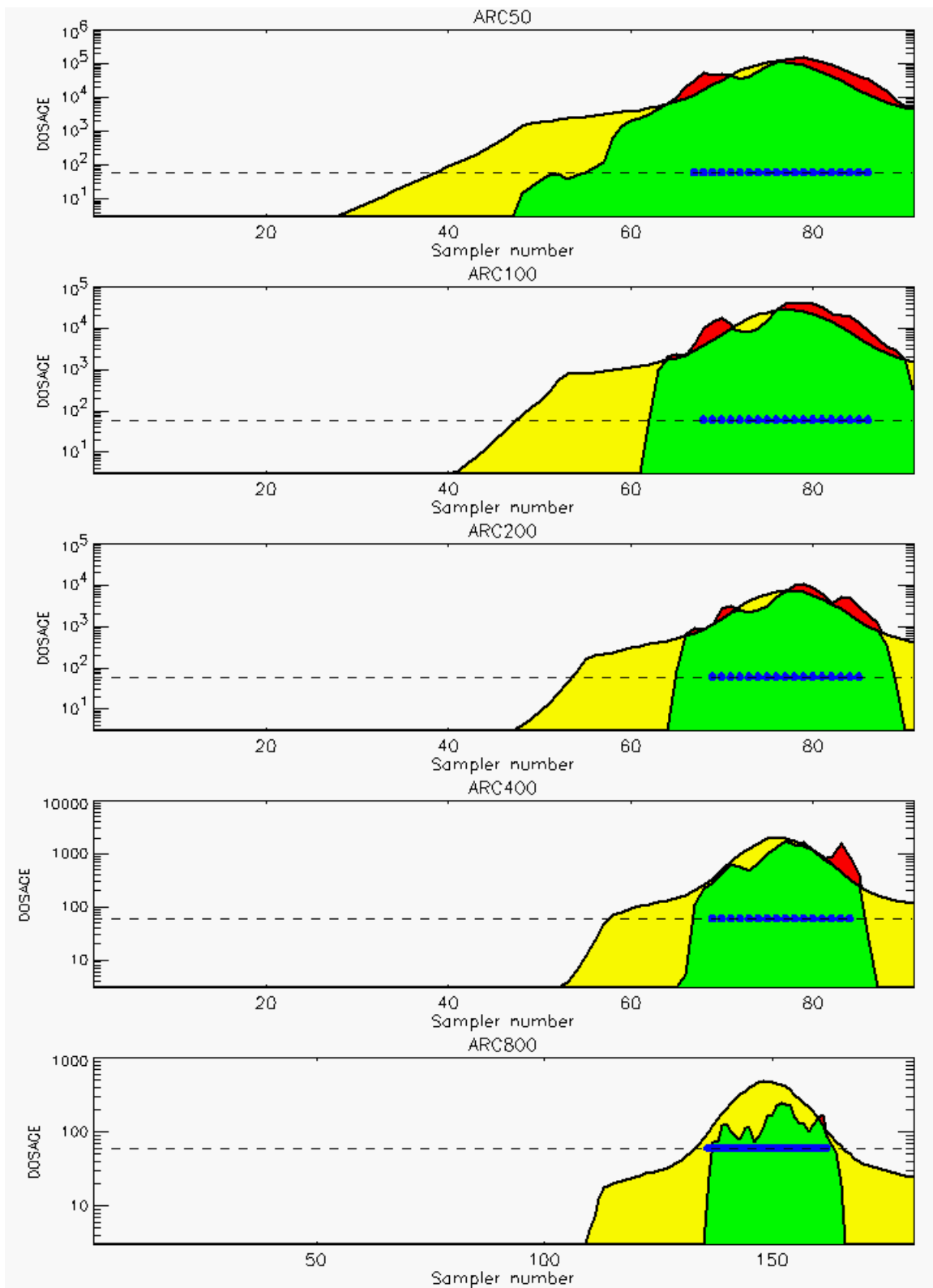


Figure D-42. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 51: Stability Class is 2

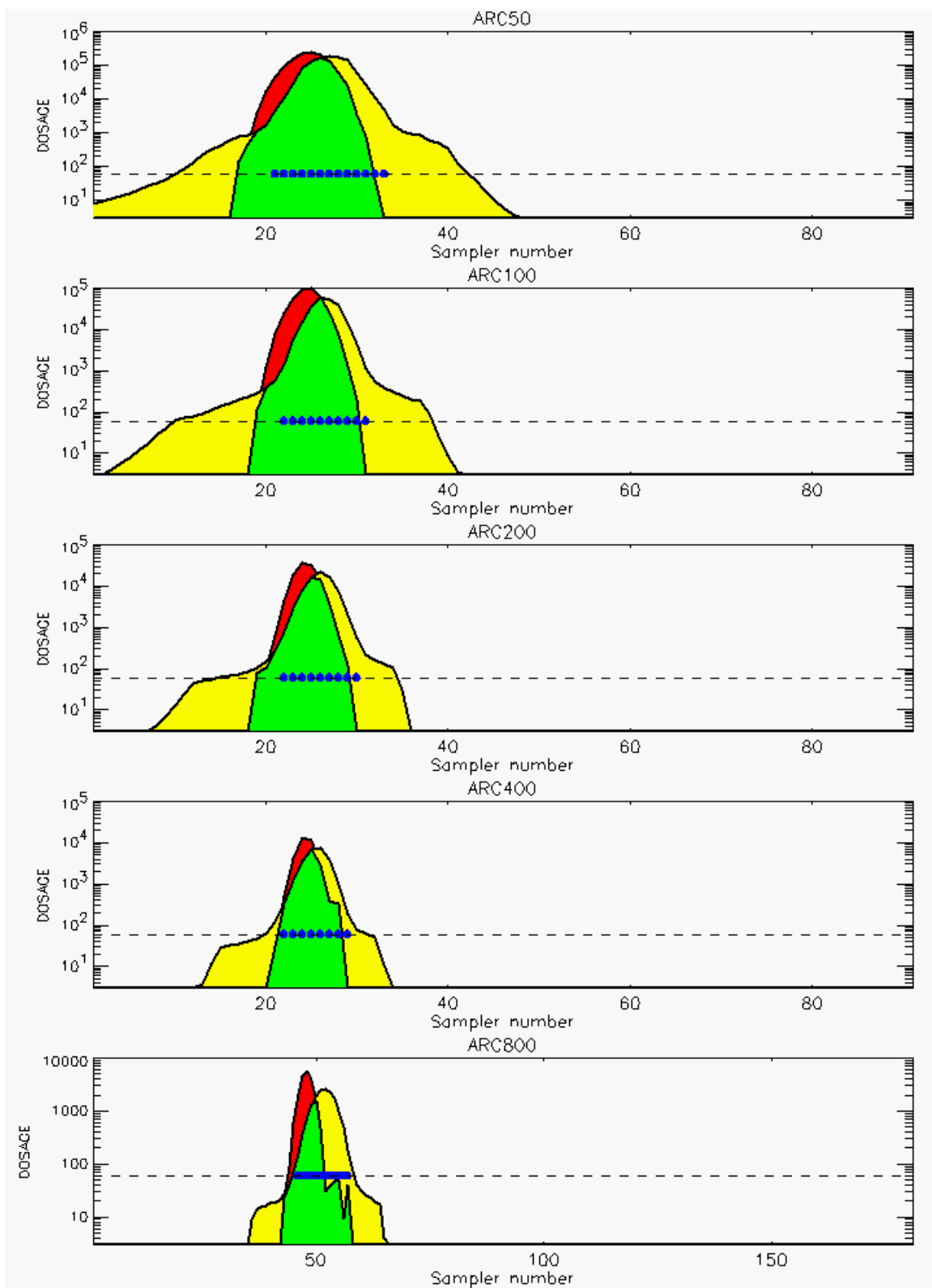


Figure D-43. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 54: Stability Class is 5

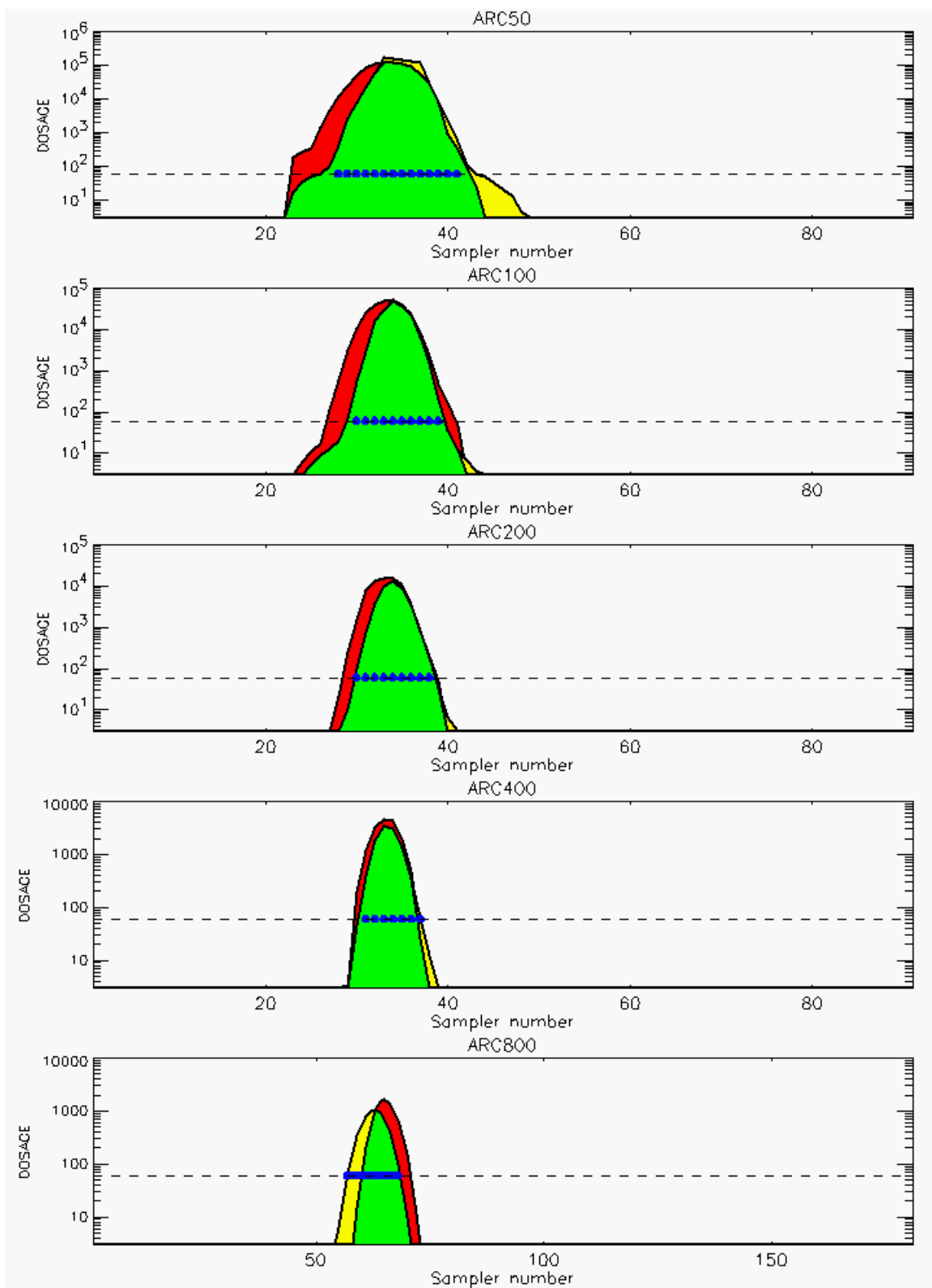


Figure D-44. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 55: Stability Class is 4

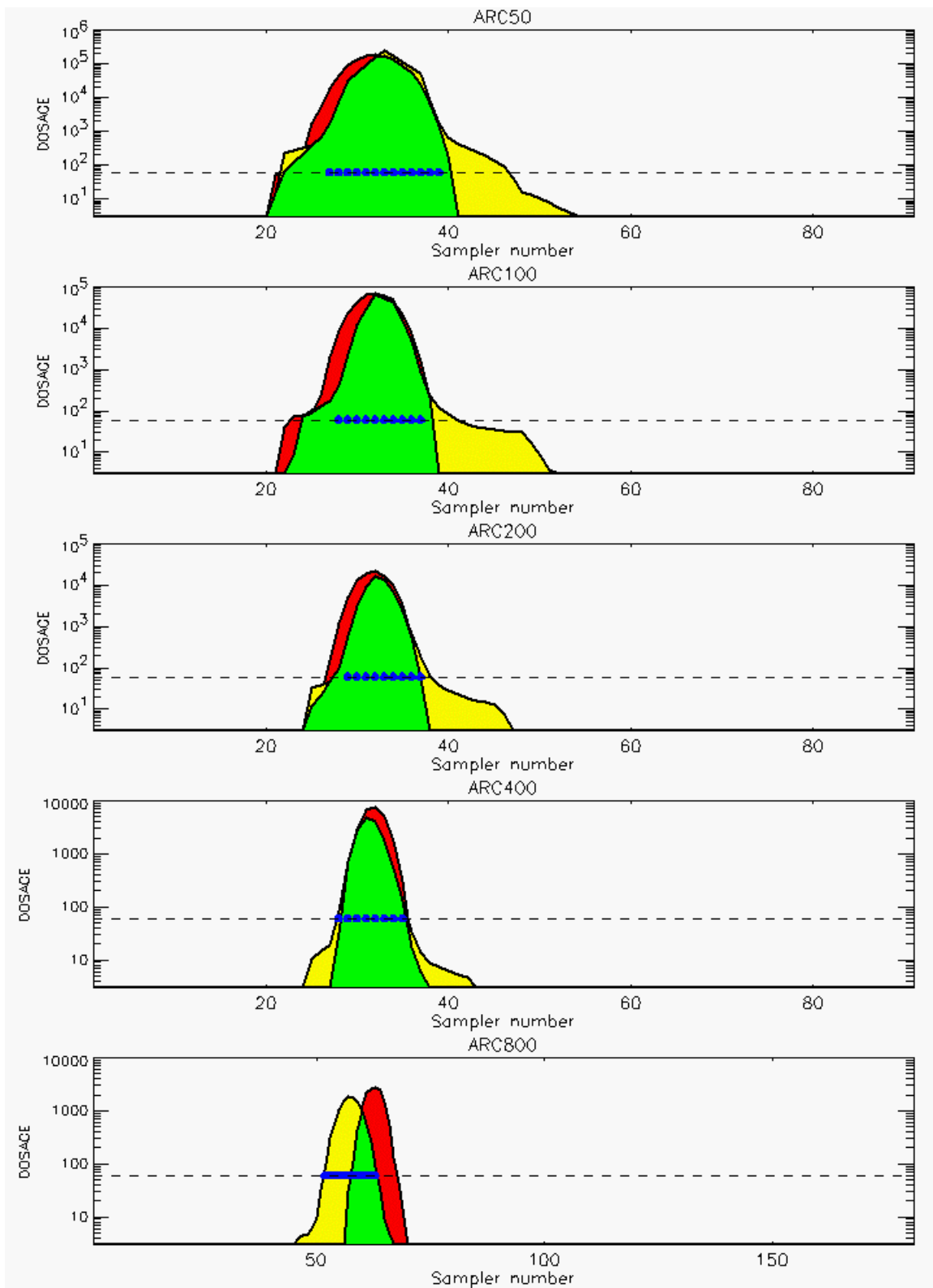


Figure D-45. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 56: Stability Class is 4

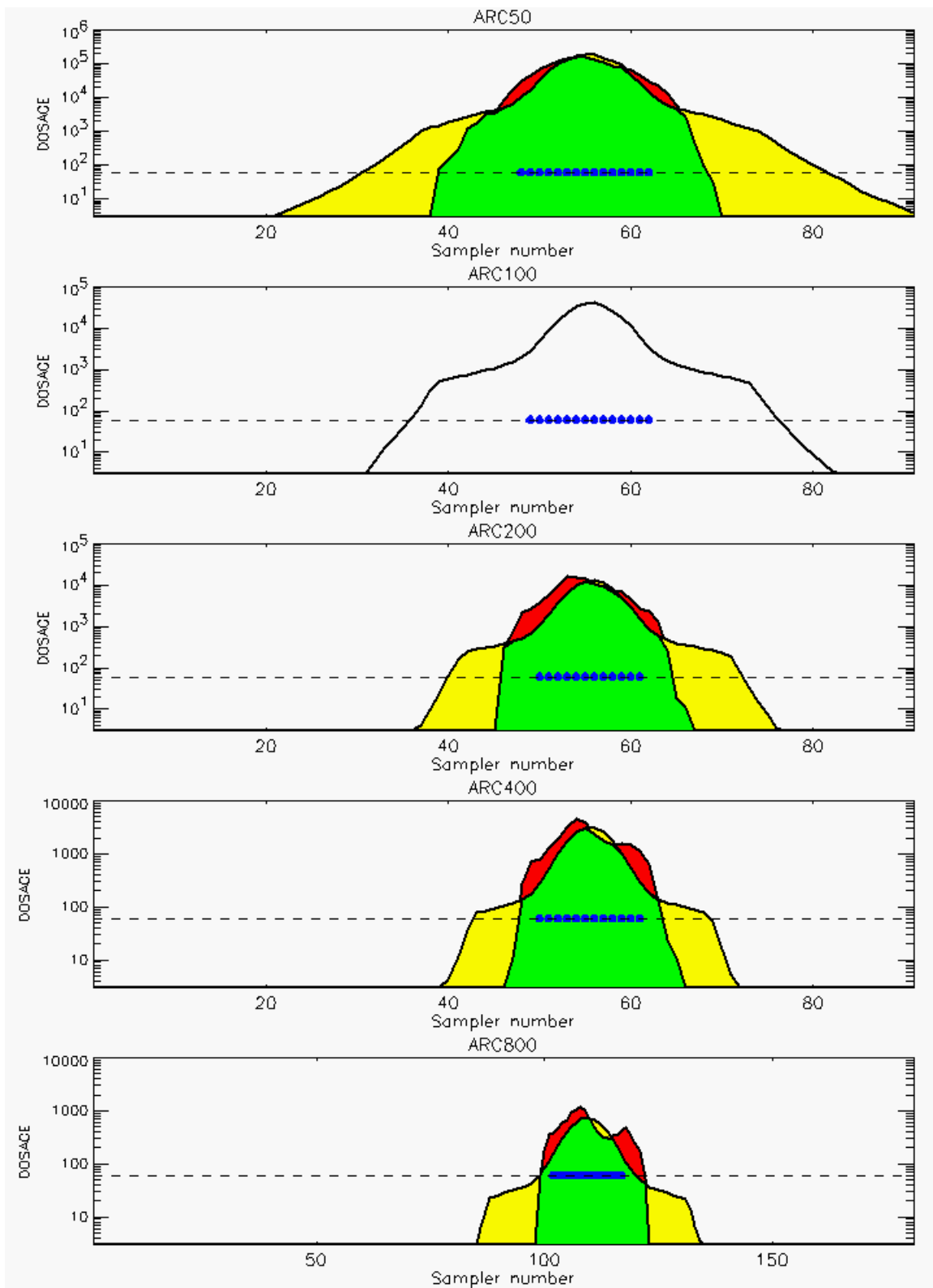


Figure D-46. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 57: Stability Class is 3

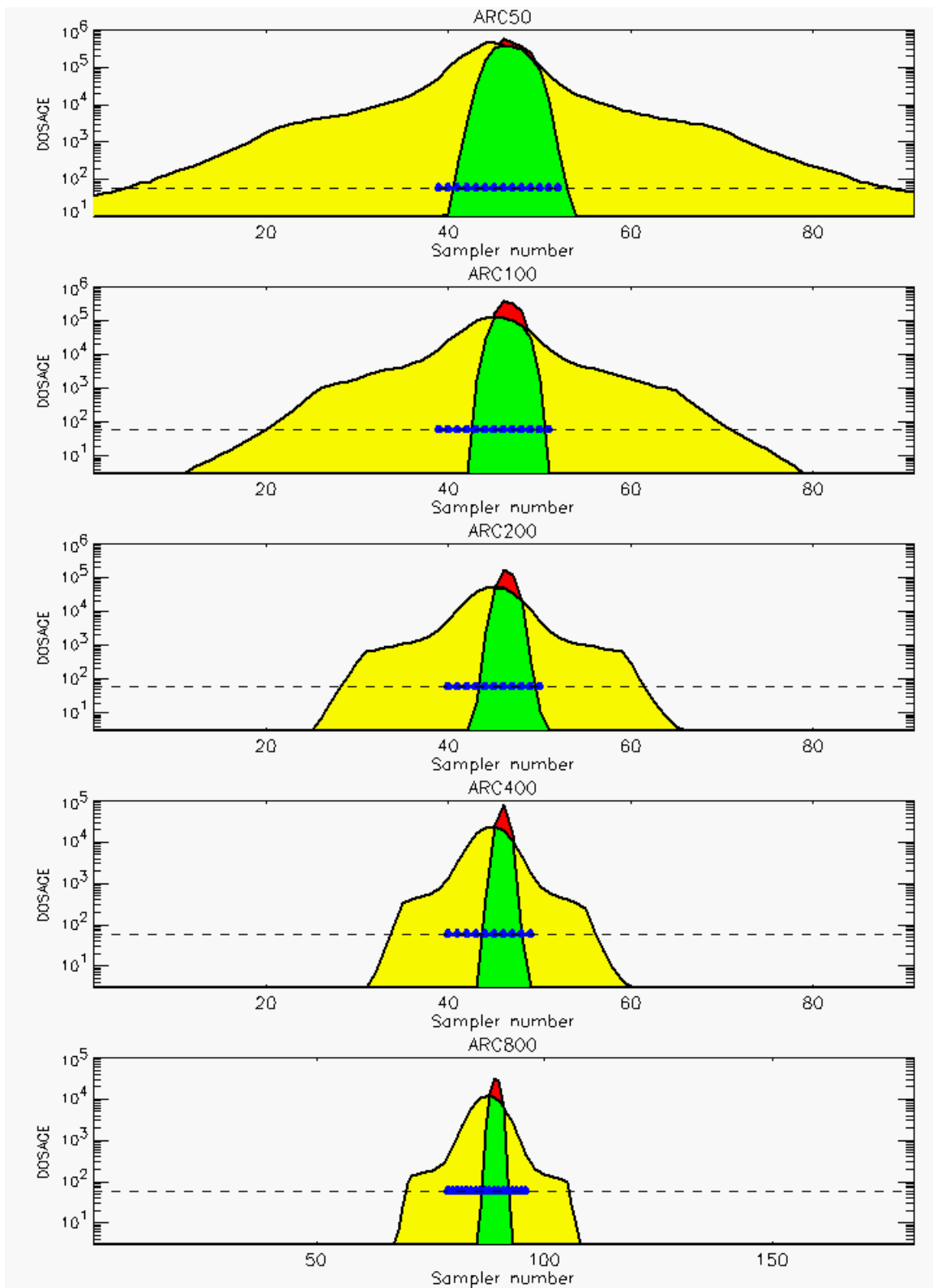


Figure D-47. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 58: Stability Class is 6

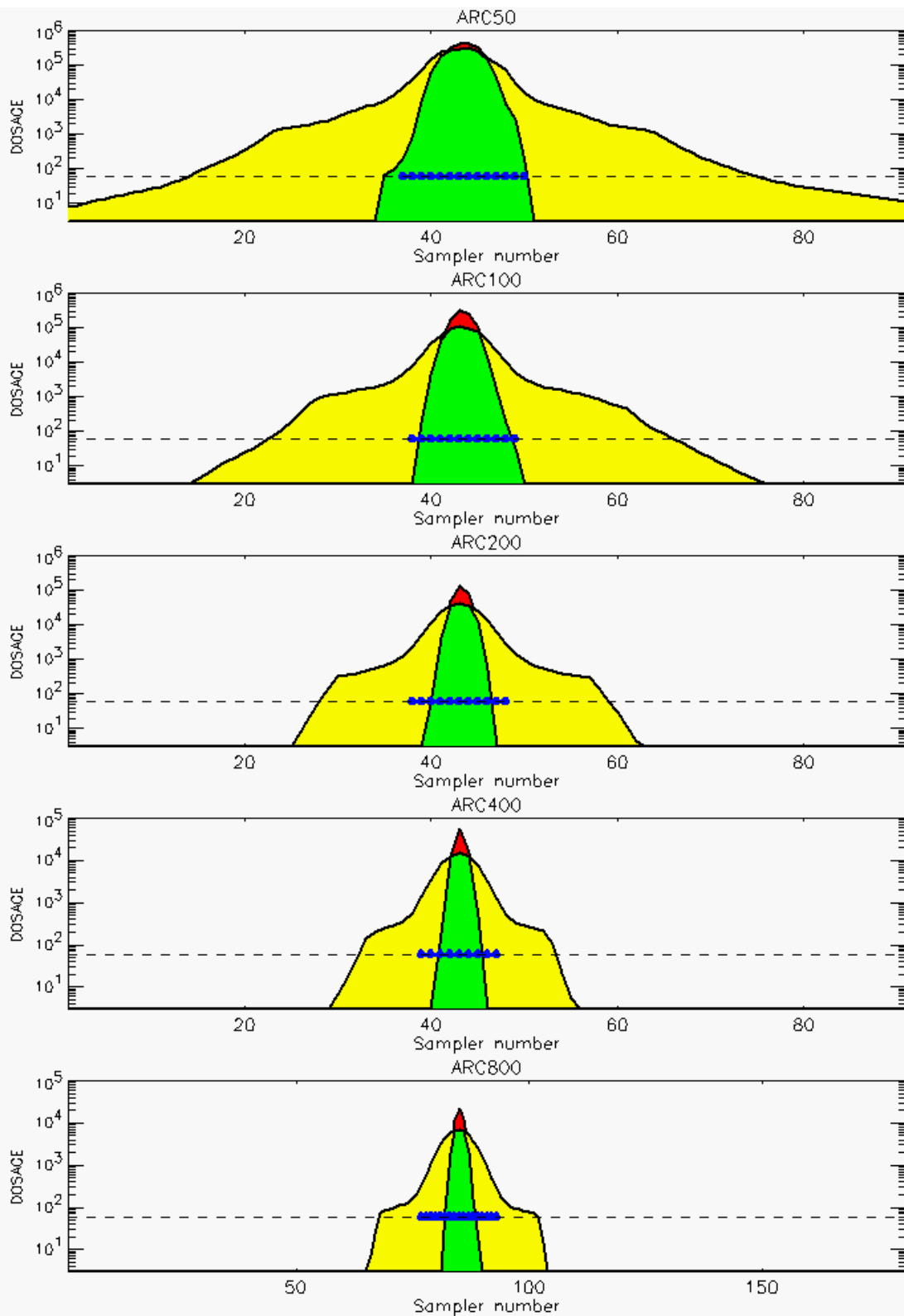


Figure D-48. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 59: Stability Class is 5

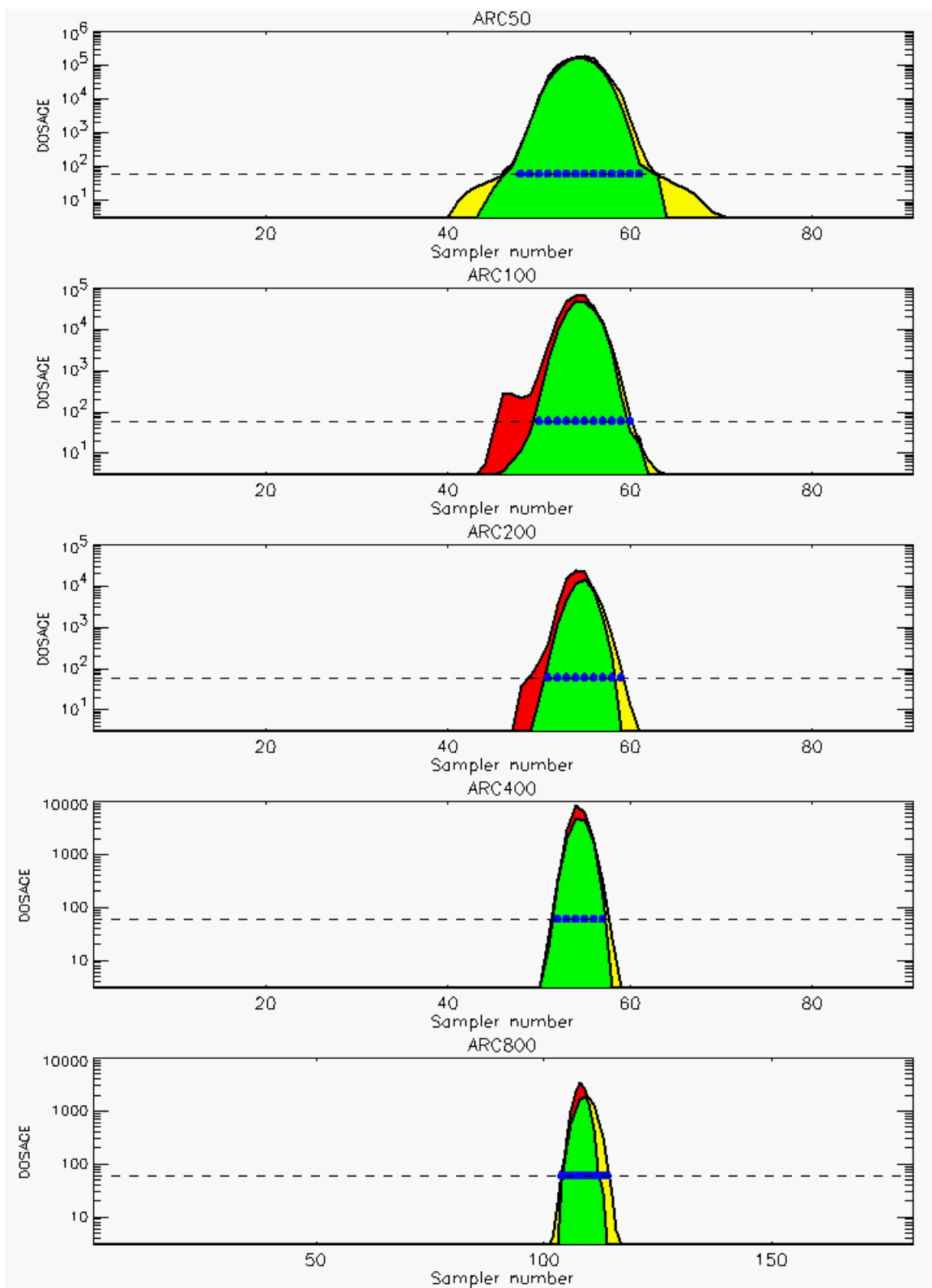


Figure D-49. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 60: Stability Class is 5

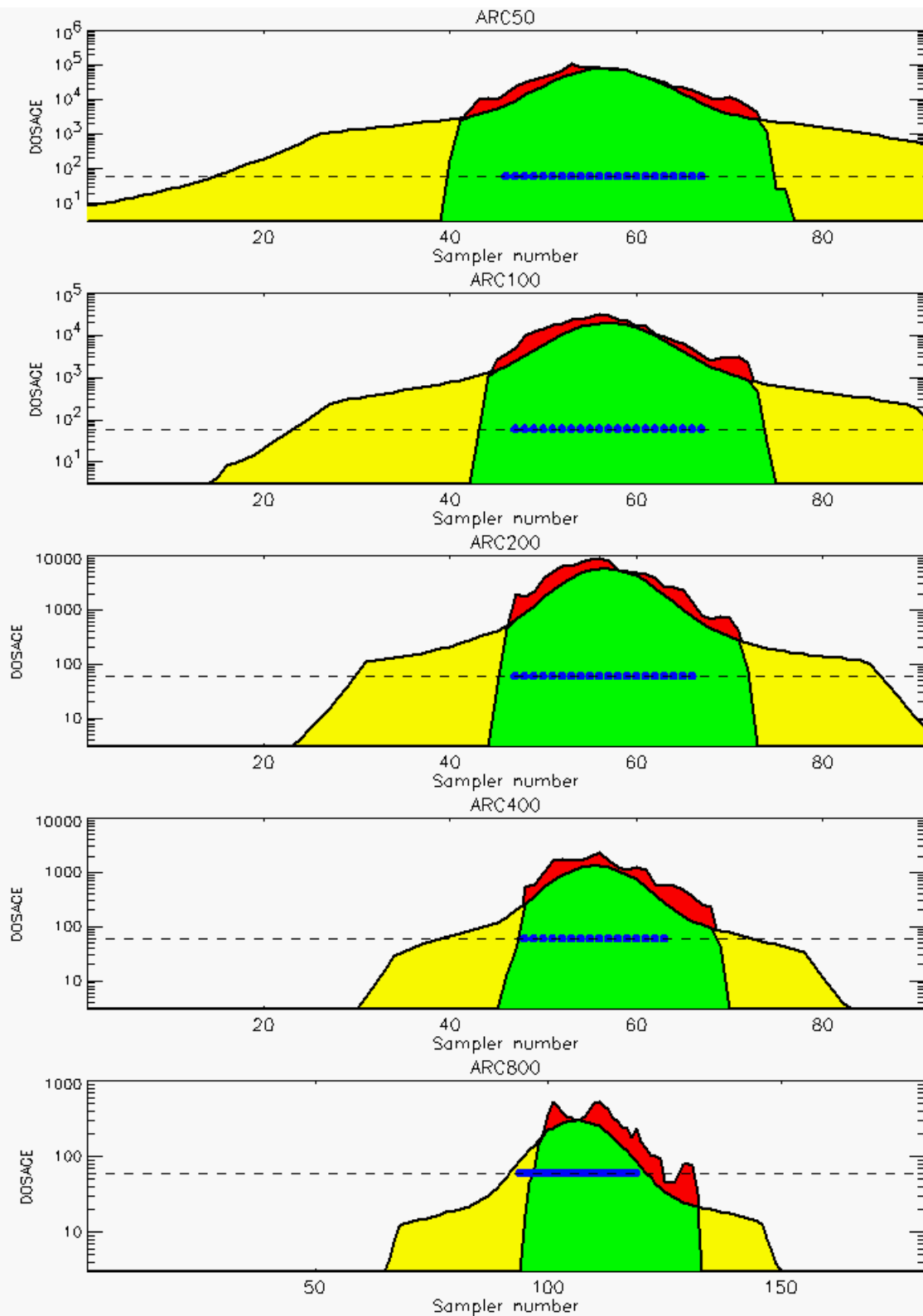


Figure D-50. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 61: Stability Class is 3

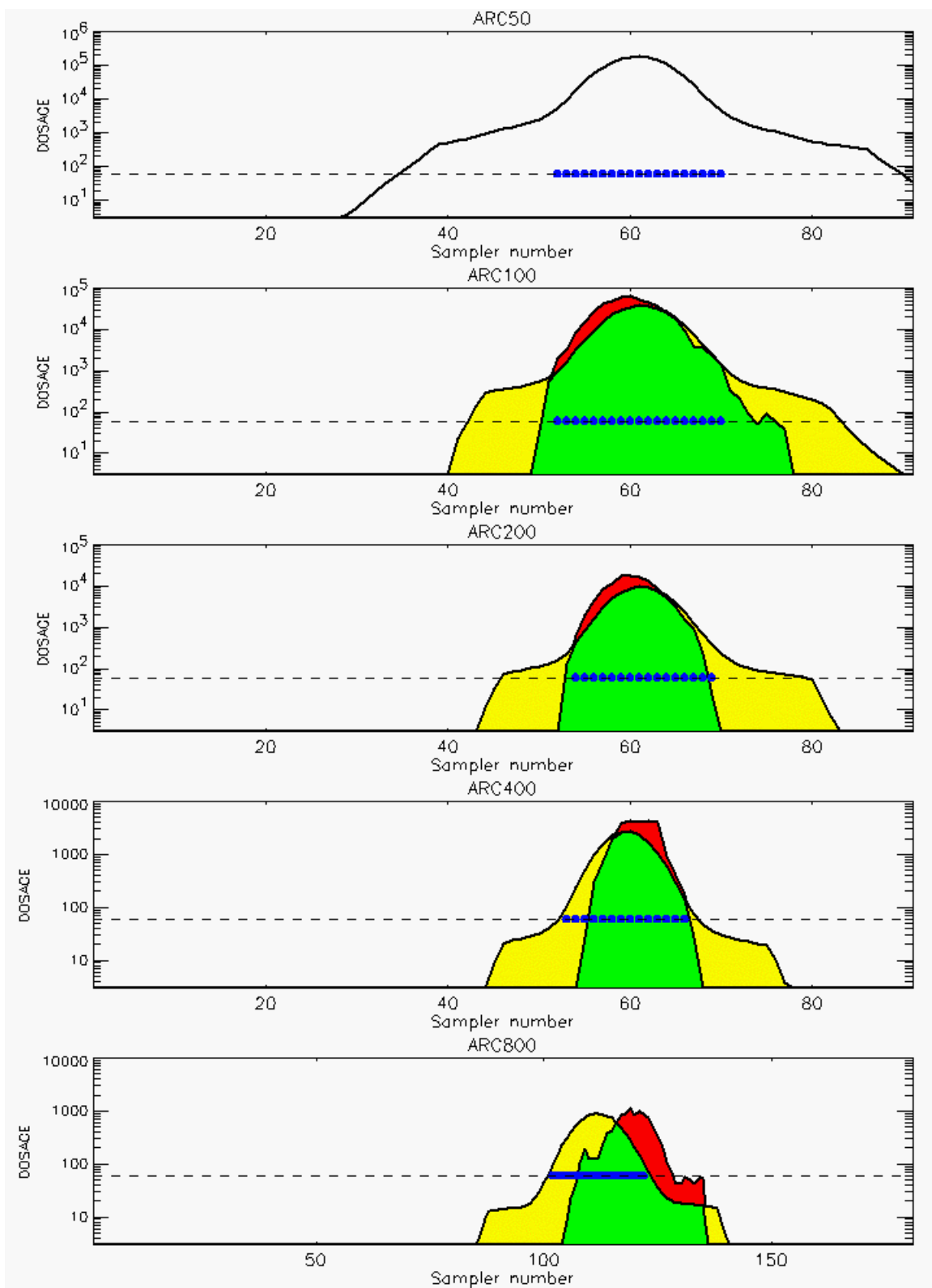


Figure D-51. 0.50 Conditional Probability Prediction and Predicted Samplers with 0.50 Probability of Exceeding Threshold Value for Trial 62: Stability Class is 2

REFERENCES

- D-1. Barad, M. L. (Editor), *Project Prairie Grass, A Field Program in Diffusion*, Geophysical Research Papers No. 59, Volumes I and II, DTIC #AD-152572/AFCRC-TR-58-235(I), Air Geophysical Laboratory, Hanscom Air Force Base, MA, 1958.
- D-2. Barad, M. L. (Editor), *Project Prairie Grass, a Field Program in Diffusion*, Geophysical Research Papers, No. 59, Volume I, DTIC #AD-152572/AFCRC-TR-58-235(I), pages 79-80, 201, 1958.
- D-3. Warner S., Platt, N., Heagy, J. F., Bradley, S., Bieberbach, G., Sugiyama, G., Nasstrom, J. S., Foster, K. T, and Larson, D., *User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models*, IDA Paper P-3554, Appendix K, January 2001.
- D-4. Irwin, J. S. and Rosu, M-R., "Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models," *Proceedings of the 10th Joint Conference on the Applications of Air Pollution Meteorology*, American Meteorological Society, Boston, pp. 6-10, 1998.

APPENDIX E
RESULTS AND ANALYSES: SUPPLEMENTAL FIGURES AND
TABLES

APPENDIX E

RESULTS AND ANALYSES: SUPPLEMENTAL FIGURES

This appendix provides 18 figures that show MOE confidence regions for various HPAC probabilistic prediction outputs of the *Prairie Grass* field trials [Ref. E-1]. Each figure displays six plots, with the exception of Figure E-3, which has only four plots. Each plot presents eight 95th percent confidence regions. The yellow region in each plot corresponds to the MOE estimate based on the standard HPAC mean value prediction. The other seven regions correspond, as shown in the legend of each plot, to the probabilistic prediction outputs of 0.01, 0.50, 0.80, 0.85, 0.90, 0.95, and 0.999. Additional details associated with these probabilistic predictions can be found in Chapter 2 of this paper.

Figures E-1 through E-7 provide results for HPAC full probabilistic prediction outputs. Figure E-1 provides plots that describe the MOE results for the full probabilistic prediction outputs based on a threshold of 60 mg-sec/m³. Figure E-1 presents results for all 51 trials and arcs, and as a function of arc range (50, 100, 200, 400, and 800 meters). Figure E-2 provides similar MOE results as a function of stability category groupings (SCGs). Stability class assignments are based on Reference E-2.

Figure E-3 presents four displays that compare MOE estimates based on two interpolation techniques. This figure considers all trials and a threshold of 60 mg-sec/m³. The top two displays use the Delaunay triangulation interpolation technique (as described in Chapter 2). For these two displays, Delaunay triangulation is used to interpolate the field trial observations and probability outputs. For the field trial observations, logarithmic interpolation is used. For the output probabilities, which vary only from 0 to 1.0, both logarithmic and linear interpolation were used (as noted in the display). The bottom two displays of Figure E-3 present similar MOE estimates based on the polar bilinear interpolation technique (as described in Chapter 2).

Figures E-4 through E-7 present results as a function of SCG for Delaunay triangulation and polar bilinear interpolation techniques.

Figures E-8 through E-18 present results based on HPAC conditional probabilistic prediction outputs. Figures E-8 and E-9 consider MOE estimates based on a threshold of

60 mg-sec/m³ for all trials, as a function of arc range, and as a function of SCG. Figures E-10 and E-11 present analogous results based on summed dosage instead of a threshold.

Figure E-12 presents results based on a threshold of 60 mg-sec/m³ and based on summed dosage for all 51 trials. Delaunay triangulation and polar bilinear interpolation are compared in these plots and for summed dosages, both a linear and logarithmic scale, was used for MOE computation (as noted in the figure).

The final six figures (Figures E-13 through E-18) examine the various interpolation combinations as a function of SCG.

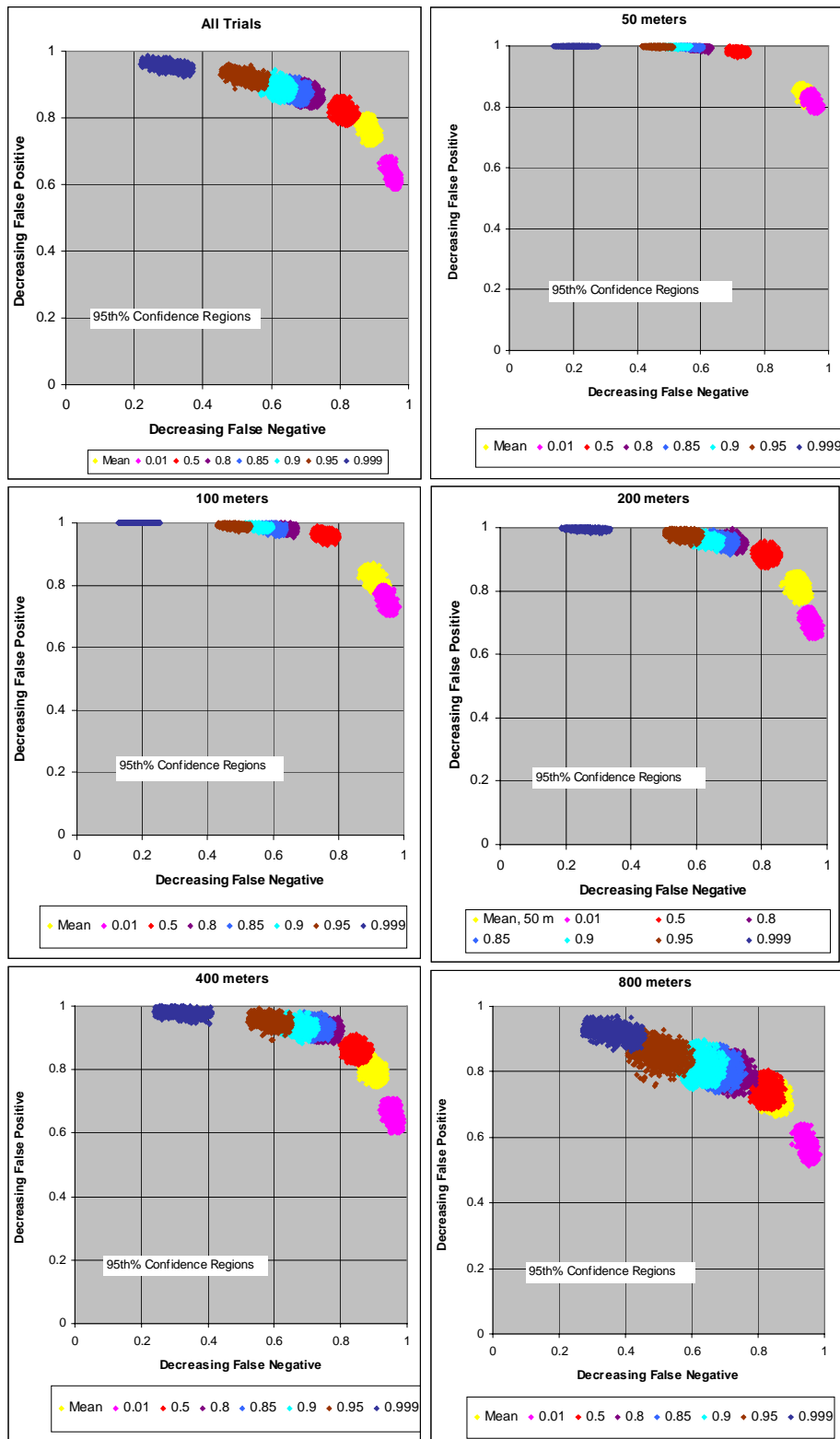


Figure E-1. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): All Trials and By Range for Threshold of 60 mg-sec/m³

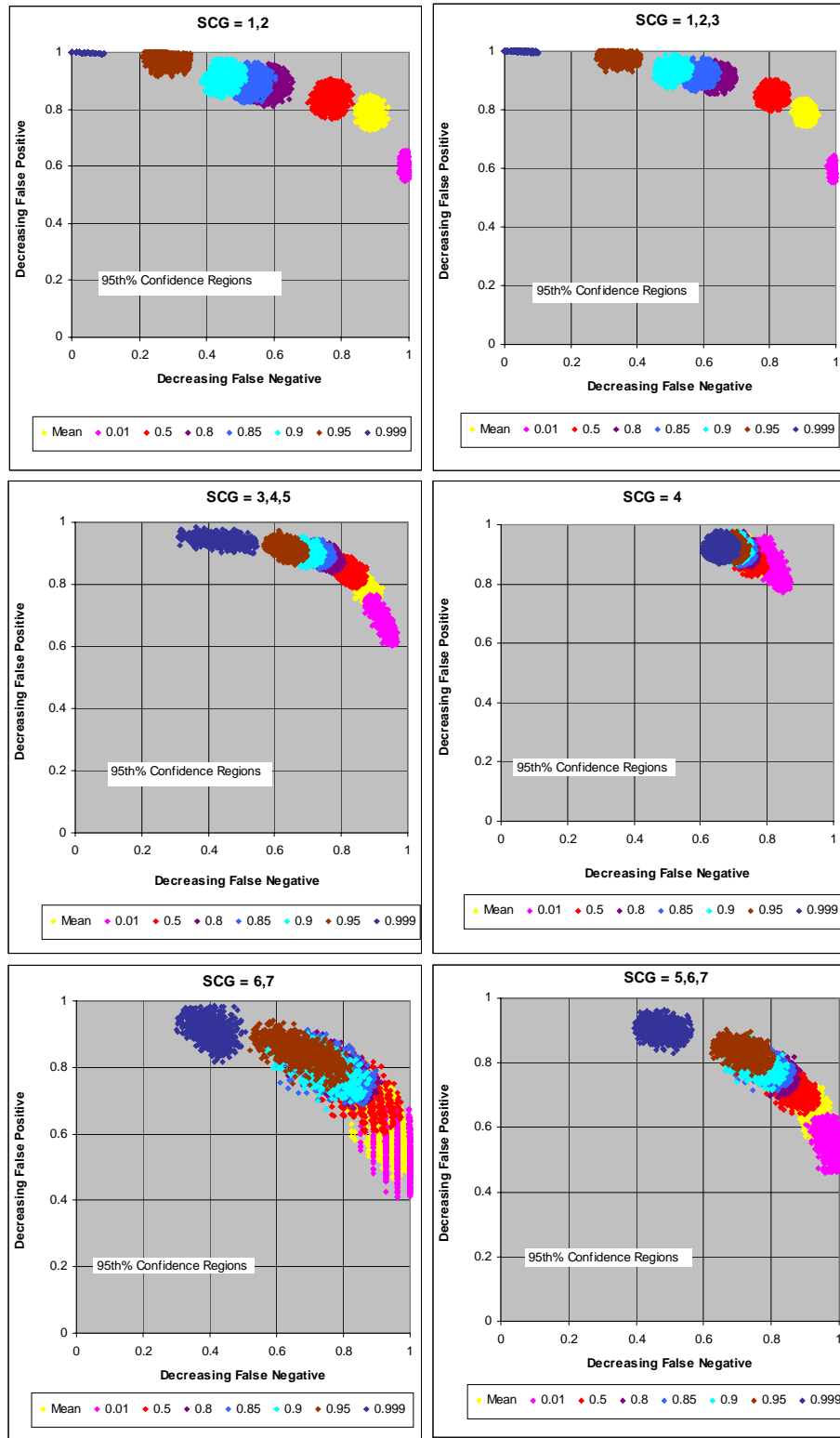


Figure E-2. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): By SCG for Threshold of 60 mg-sec/m³

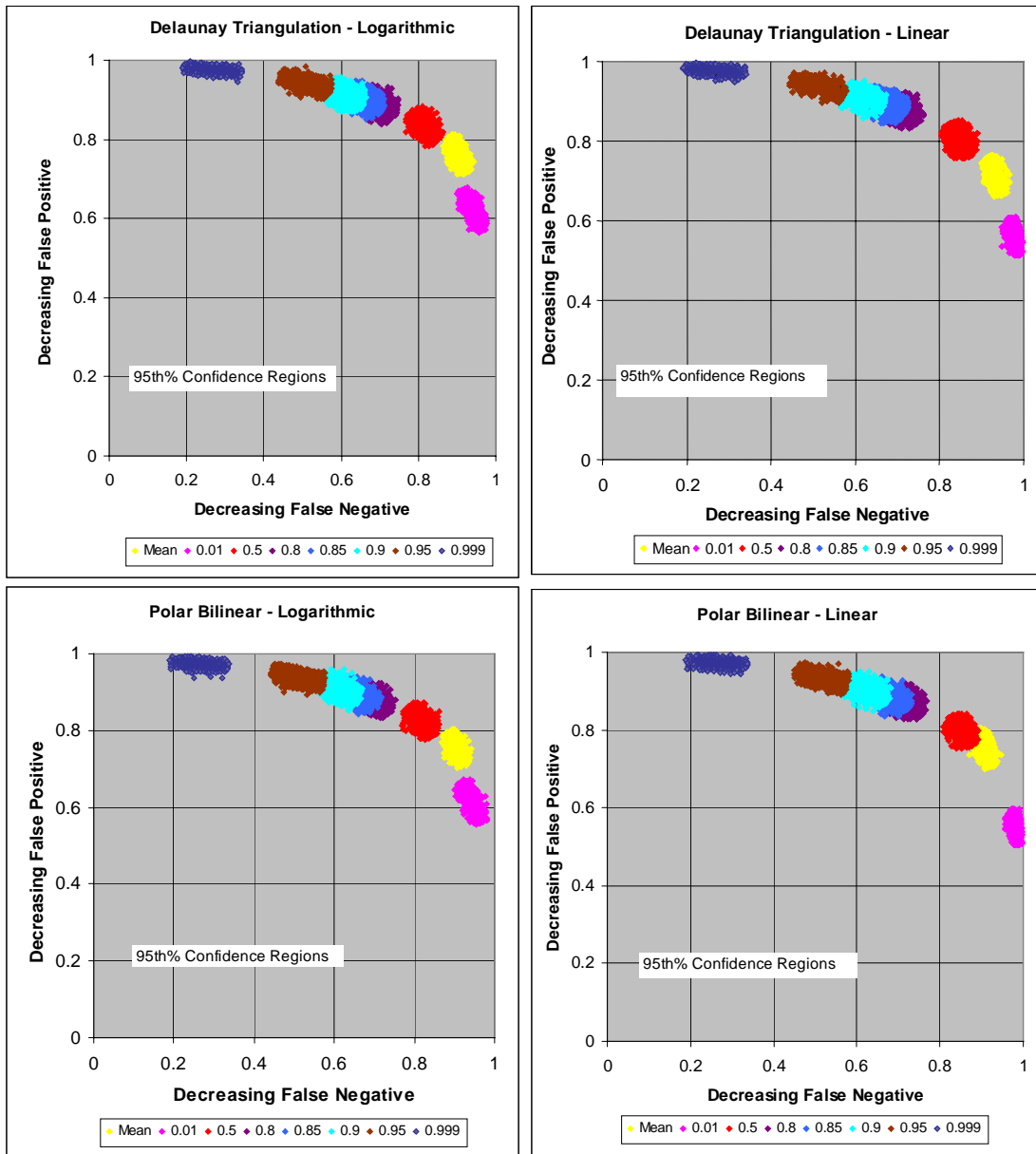


Figure E-3. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): All Trials for Threshold of 60 mg-sec/m³ and Based on Delaunay Triangulation and Polar Bilinear Interpolation (Logarithmic) of Field Trials and Probabilities (Logarithmic and Linear)

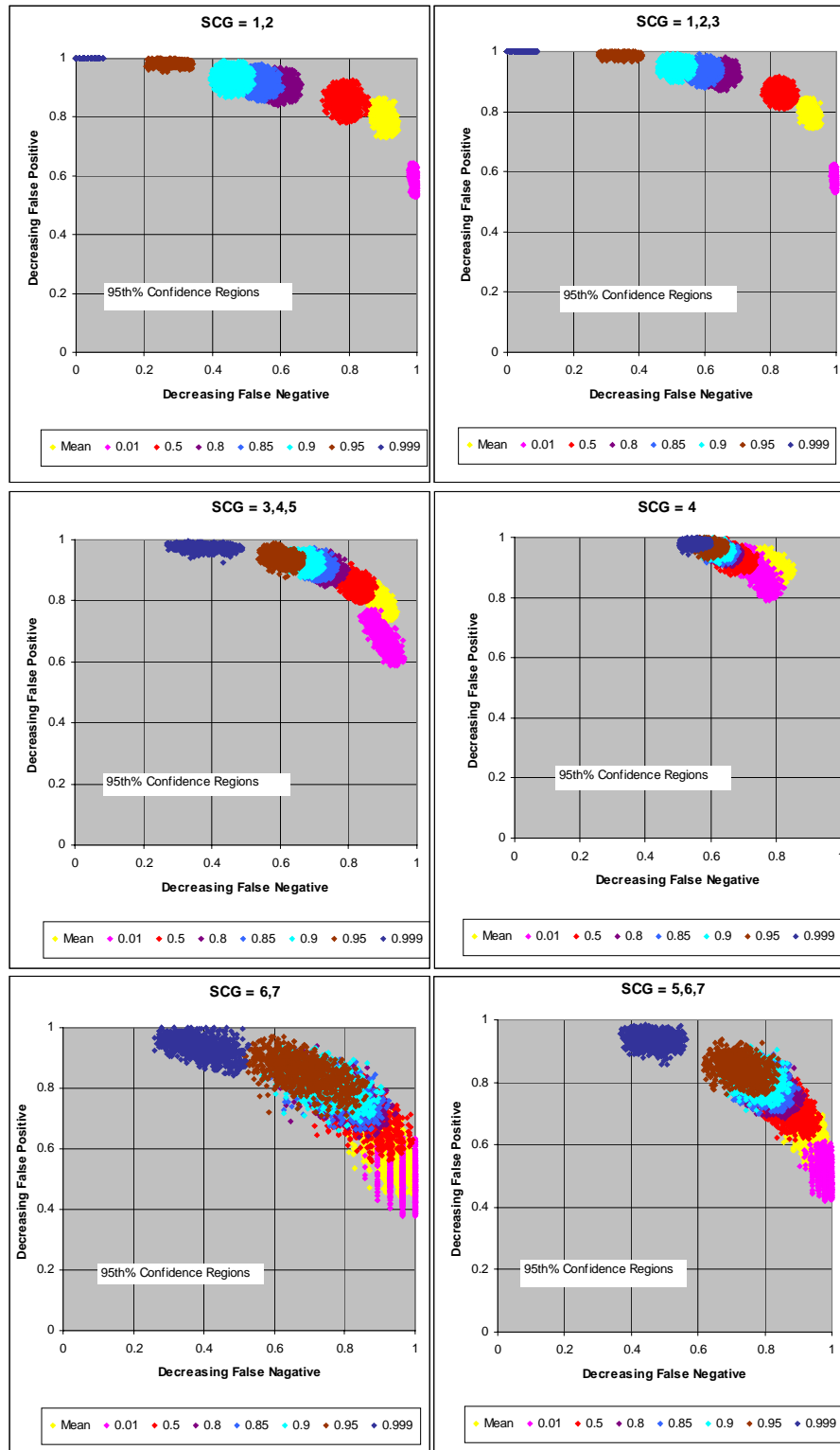


Figure E-4. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): By SCG for Threshold of 60 mg-sec/m³ and Based on Delaunay Triangulation Interpolation (Logarithmic) of Field Trials and Probabilities (Logarithmic)

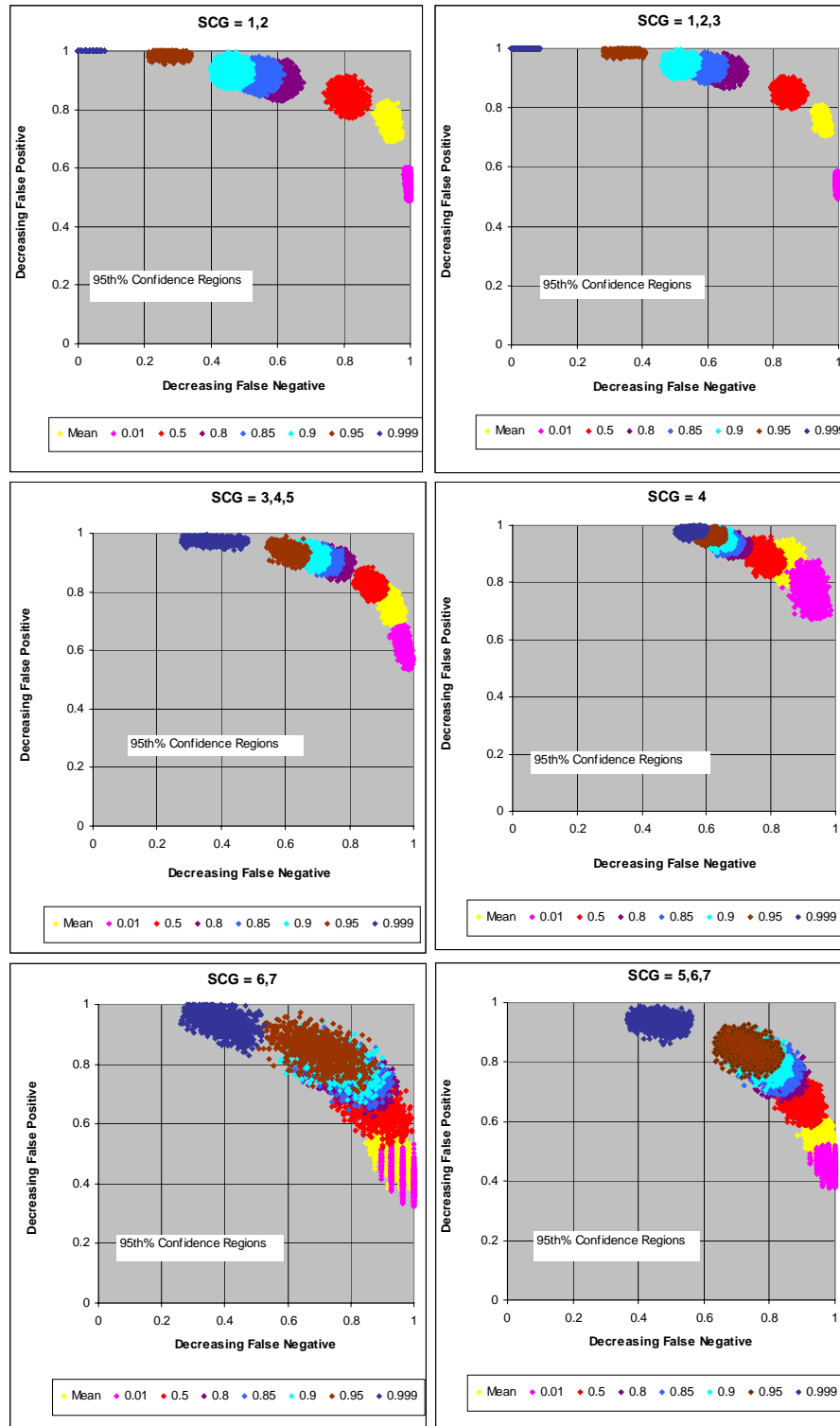


Figure E-5. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): **By SCG for Threshold of 60 mg-sec/m³ and Based on Delaunay Triangulation Interpolation of Field Trials (Logarithmic) and Probabilities (Linear)**

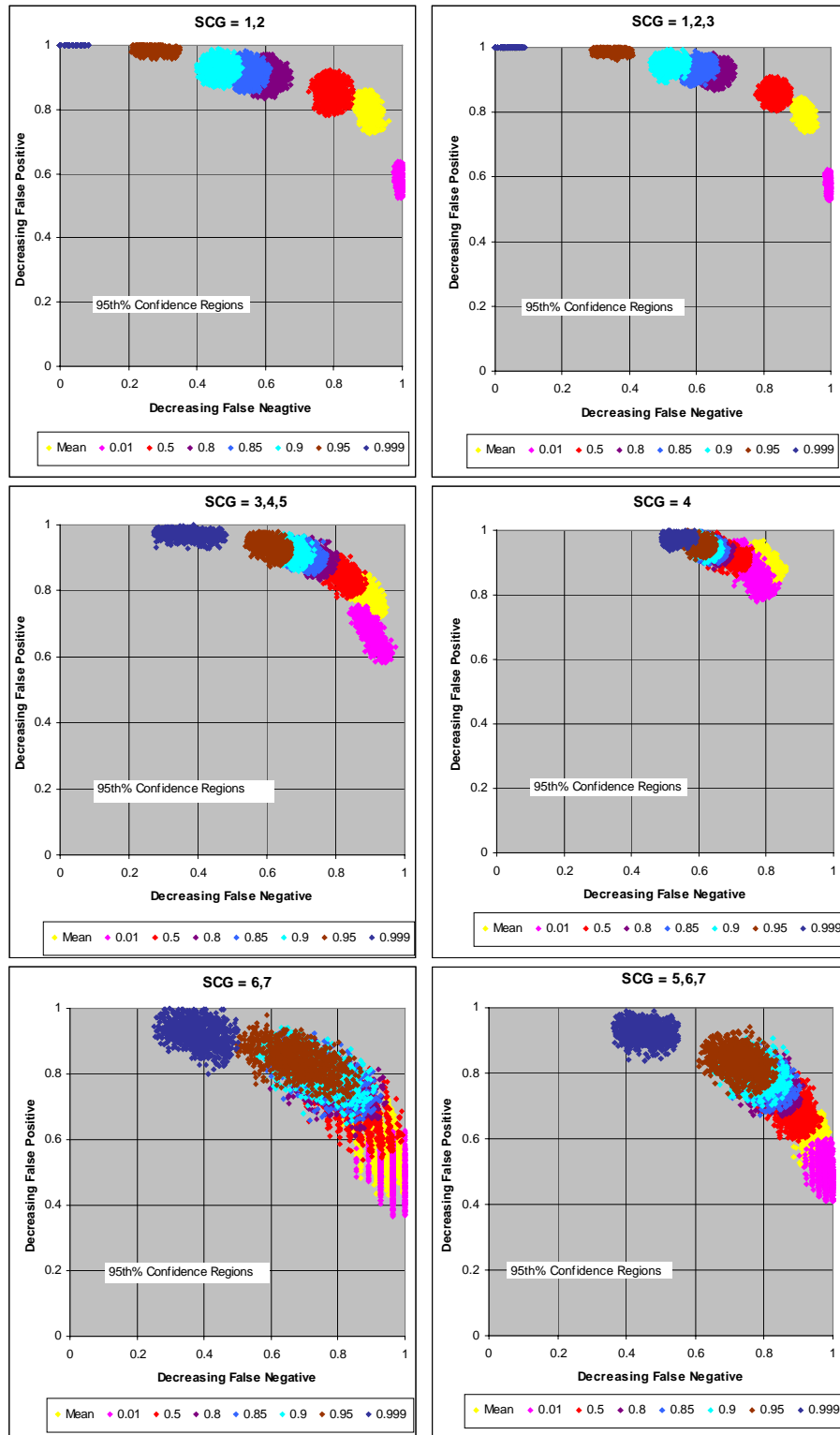


Figure E-6. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): By SCG for Threshold of 60 mg-sec/m³ and Based on Polar Bilinear Interpolation (Logarithmic) of Field Trials and Probabilities (Logarithmic)

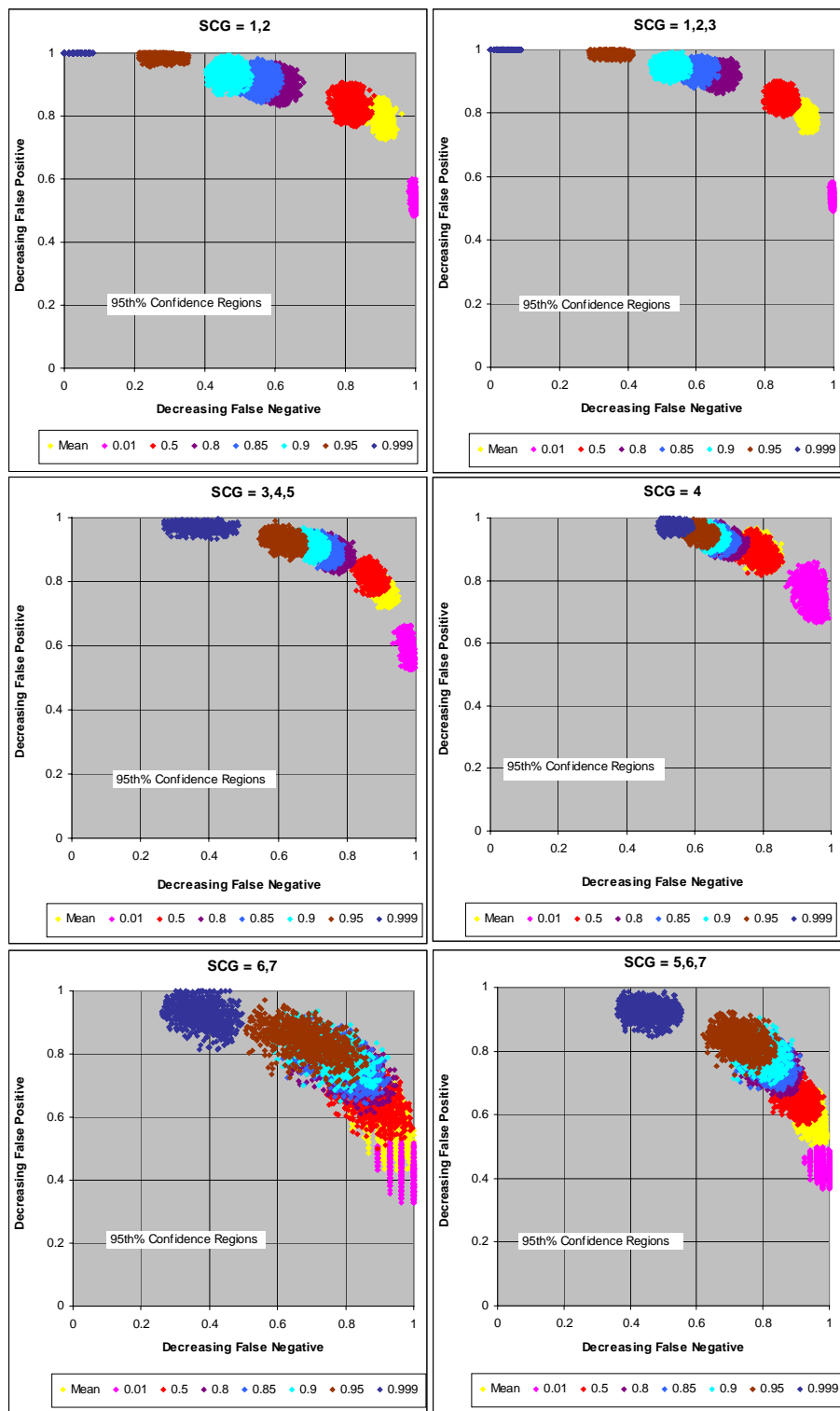


Figure E-7. MOE Confidence Regions for HPAC Full Probabilistic Prediction Outputs (and Mean Value): **By SCG for Threshold of 60 mg-sec/m³ and Based on Polar Bilinear Interpolation (Logarithmic) of Field Trials and Probabilities (Linear)**

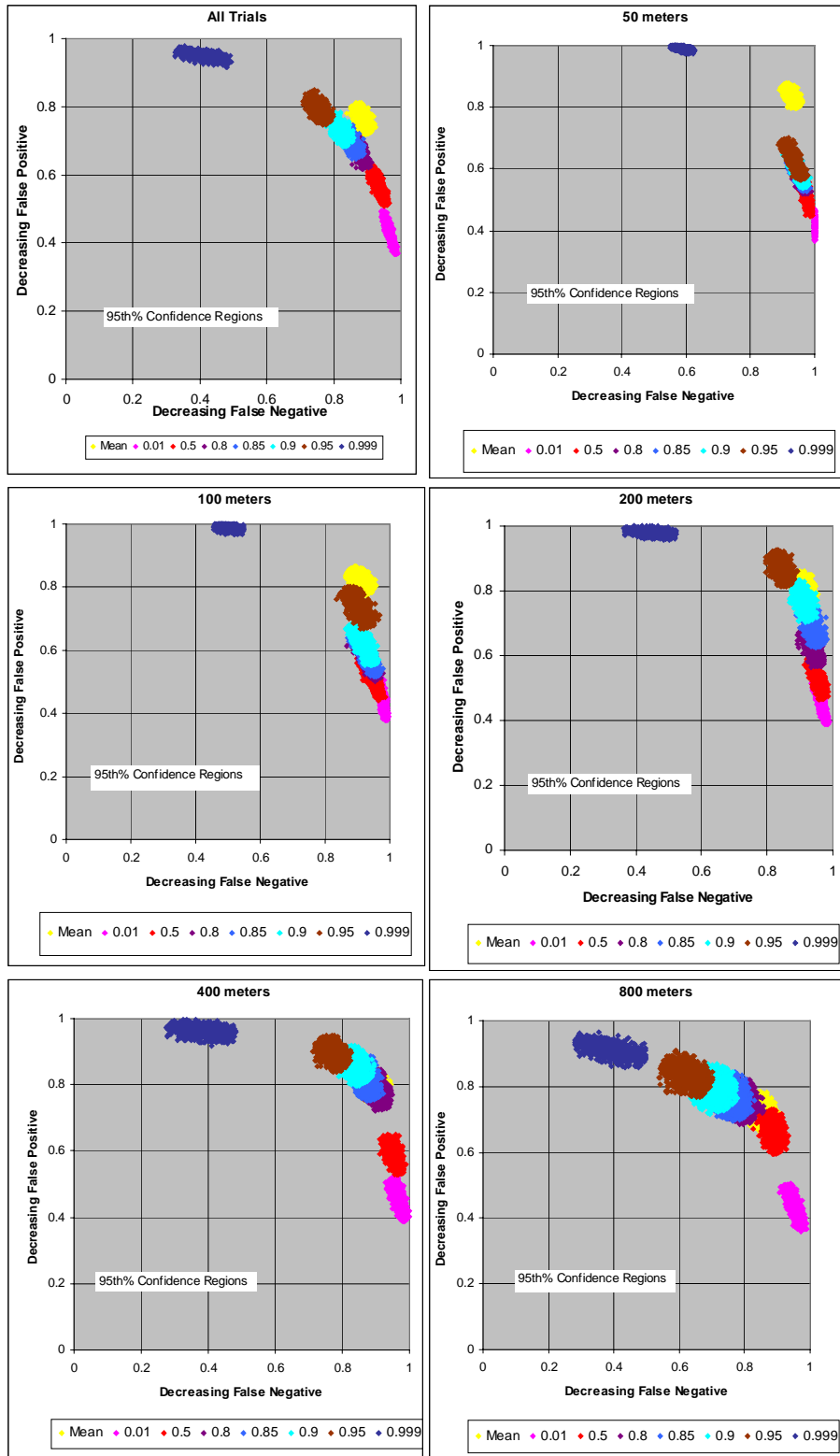


Figure E-8. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): All Trials and By Range for Threshold of 60 mg-sec/m³

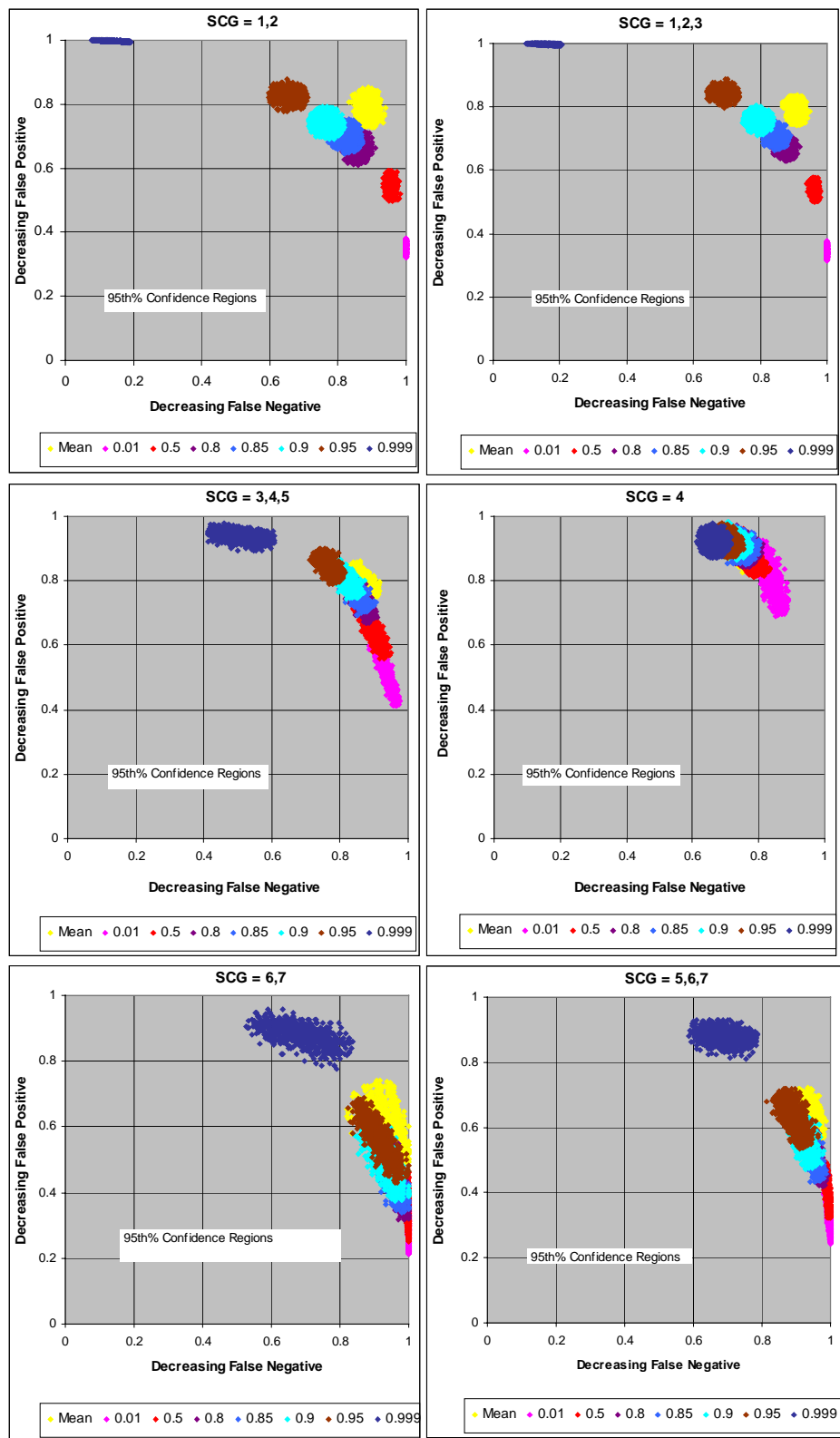


Figure E-9. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Threshold of 60 mg-sec/m³

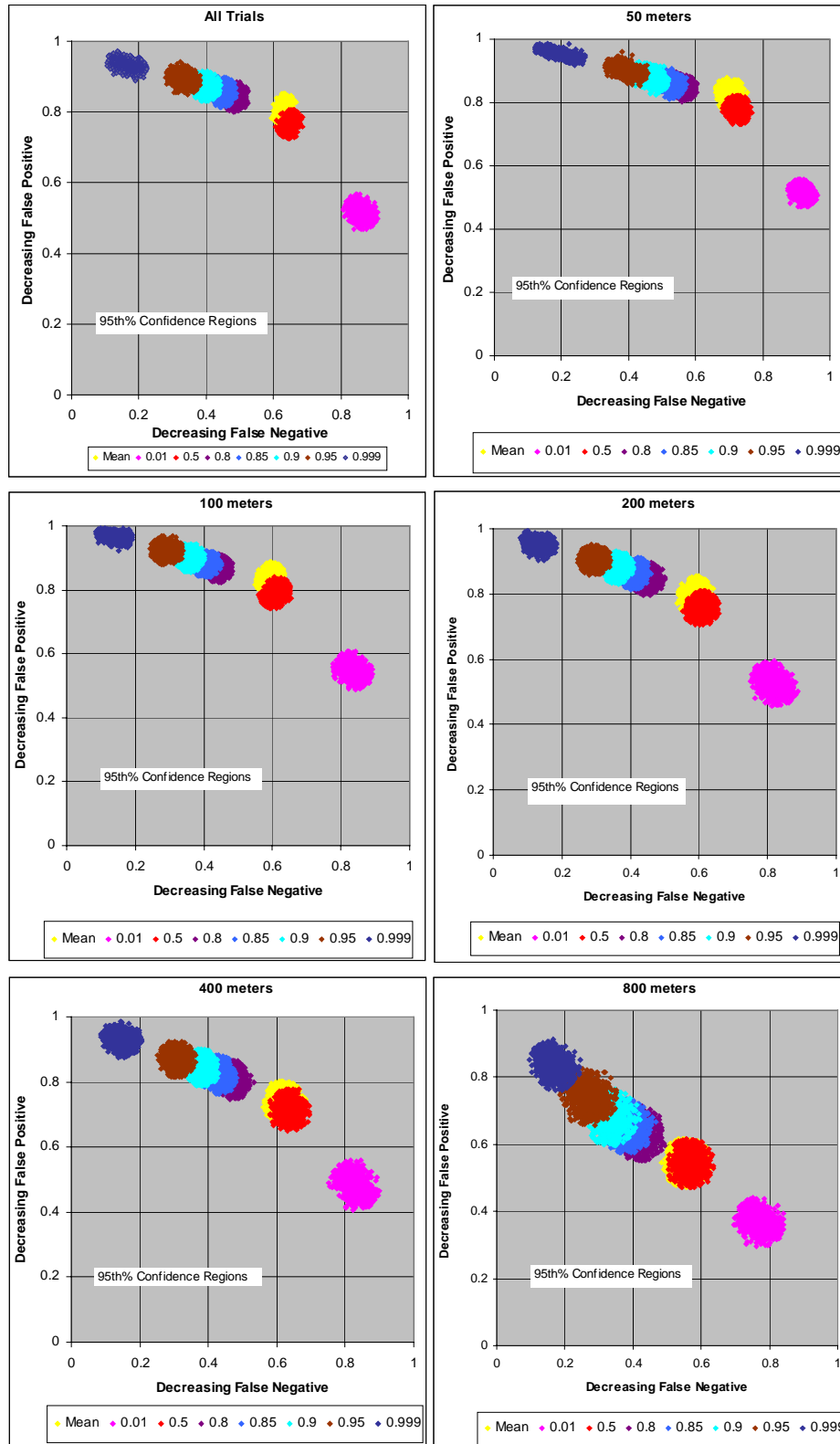


Figure E-10. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): All Trials and By Range for Summed Dosages

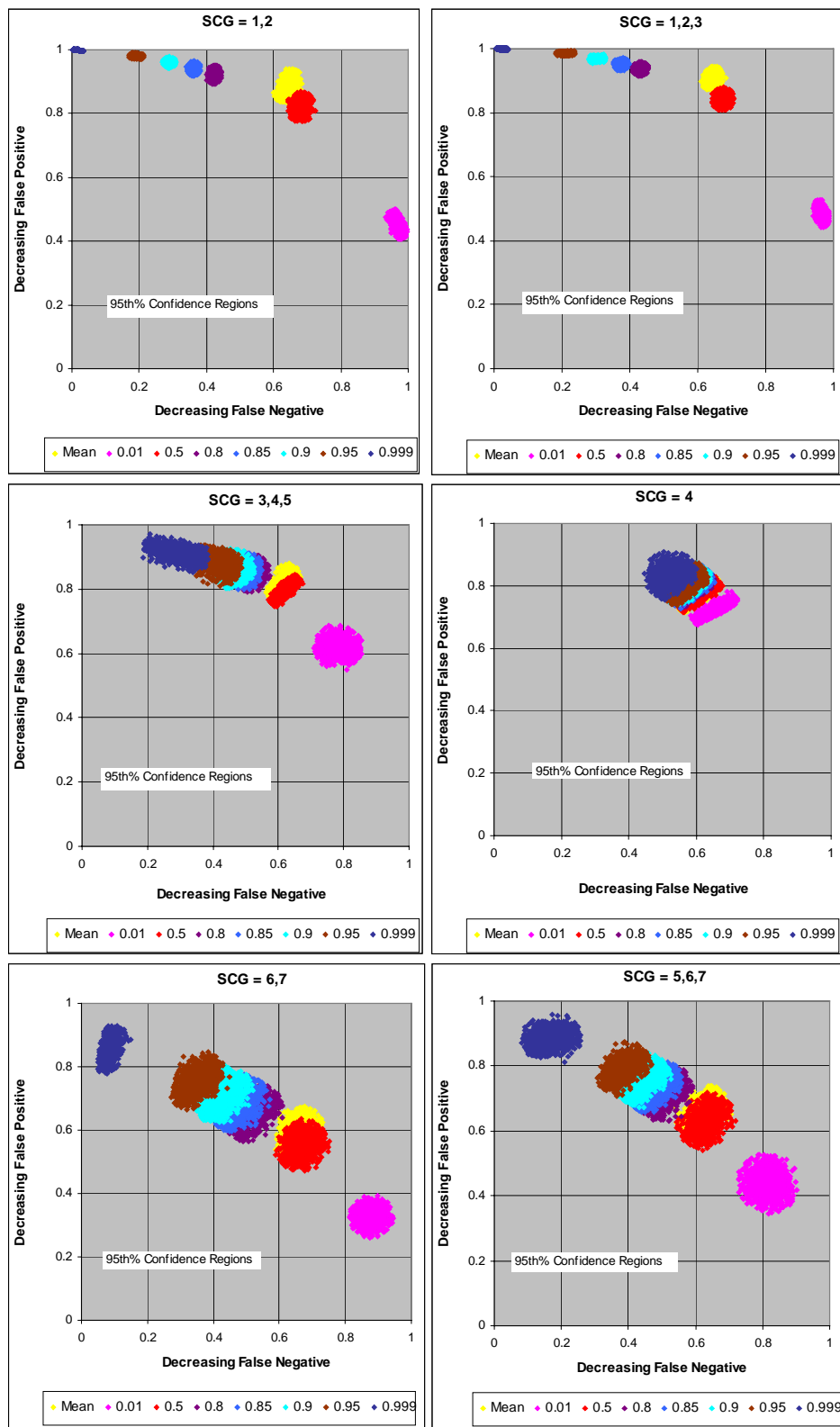


Figure E-11. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Summed Dosages

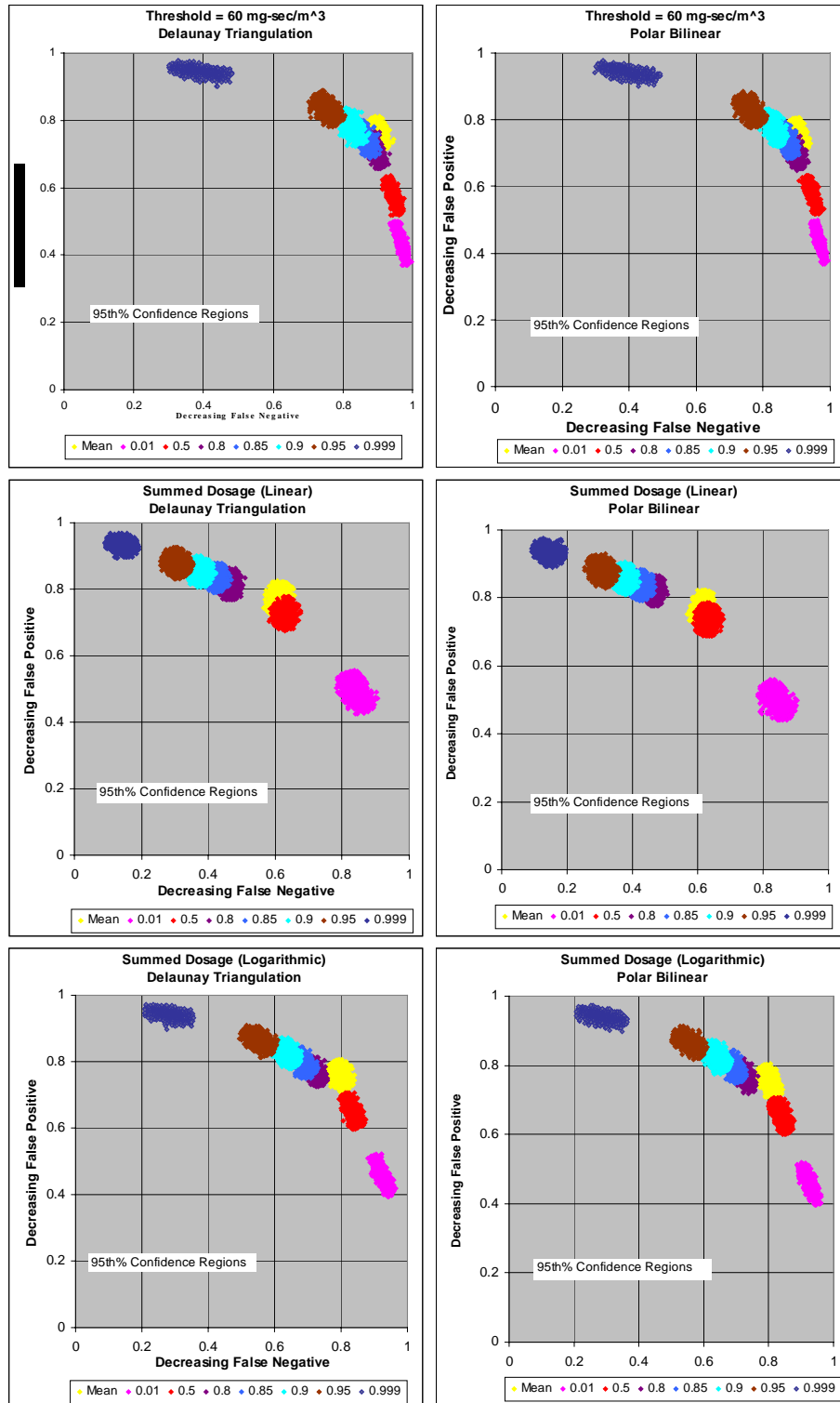


Figure E-12. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): All Trials for Threshold of 60 mg-sec/m³ and Summed Dosages Based on Delaunay Triangulation and Polar Bilinear Interpolation of Field Trials (Summed Dosage Interpolation Based on Linear and Logarithmic Scales as Noted)

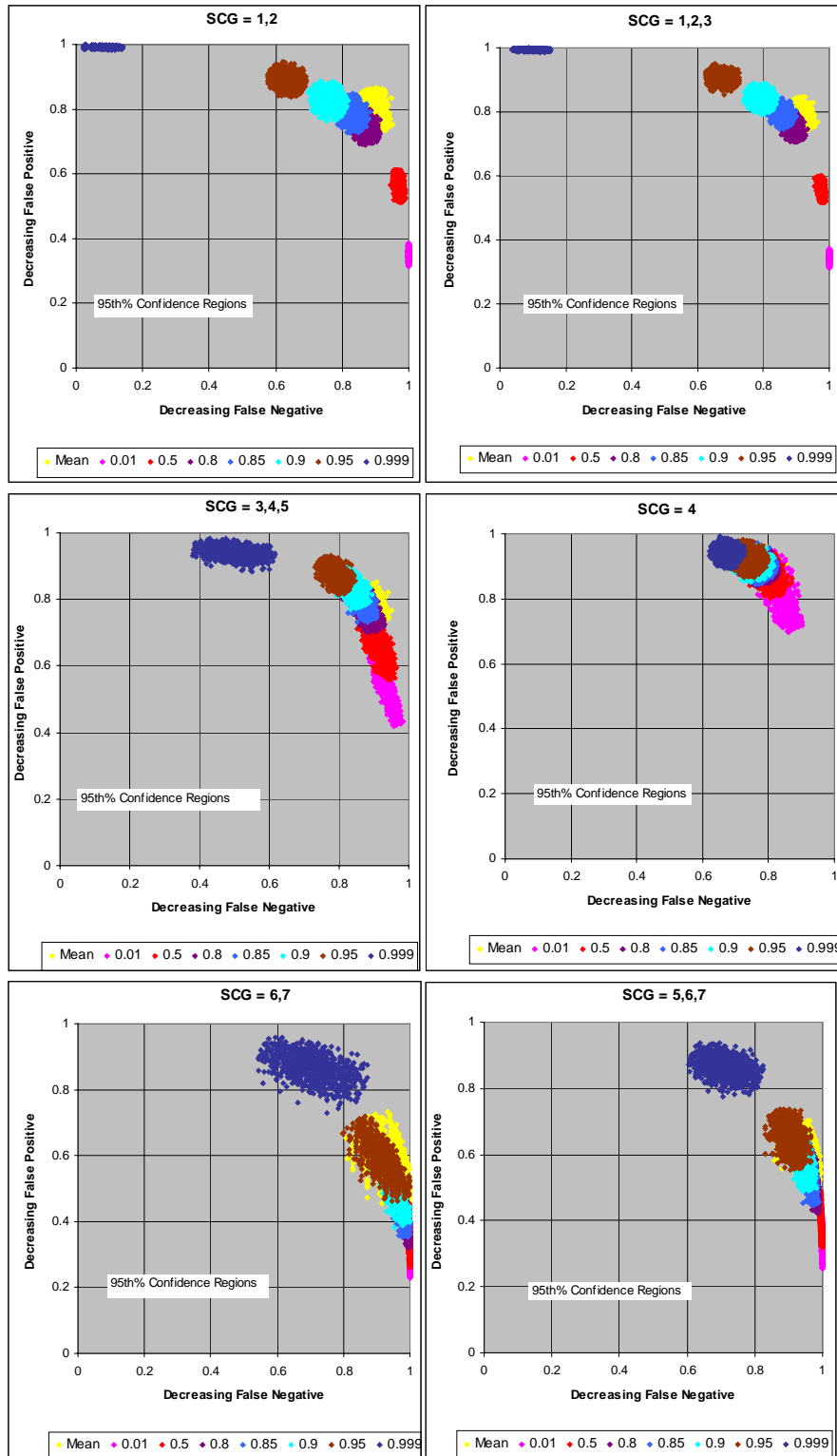


Figure E-13. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Threshold of 60 mg-sec/m³ and Based on Delaunay Triangulation Interpolation of Field Trials

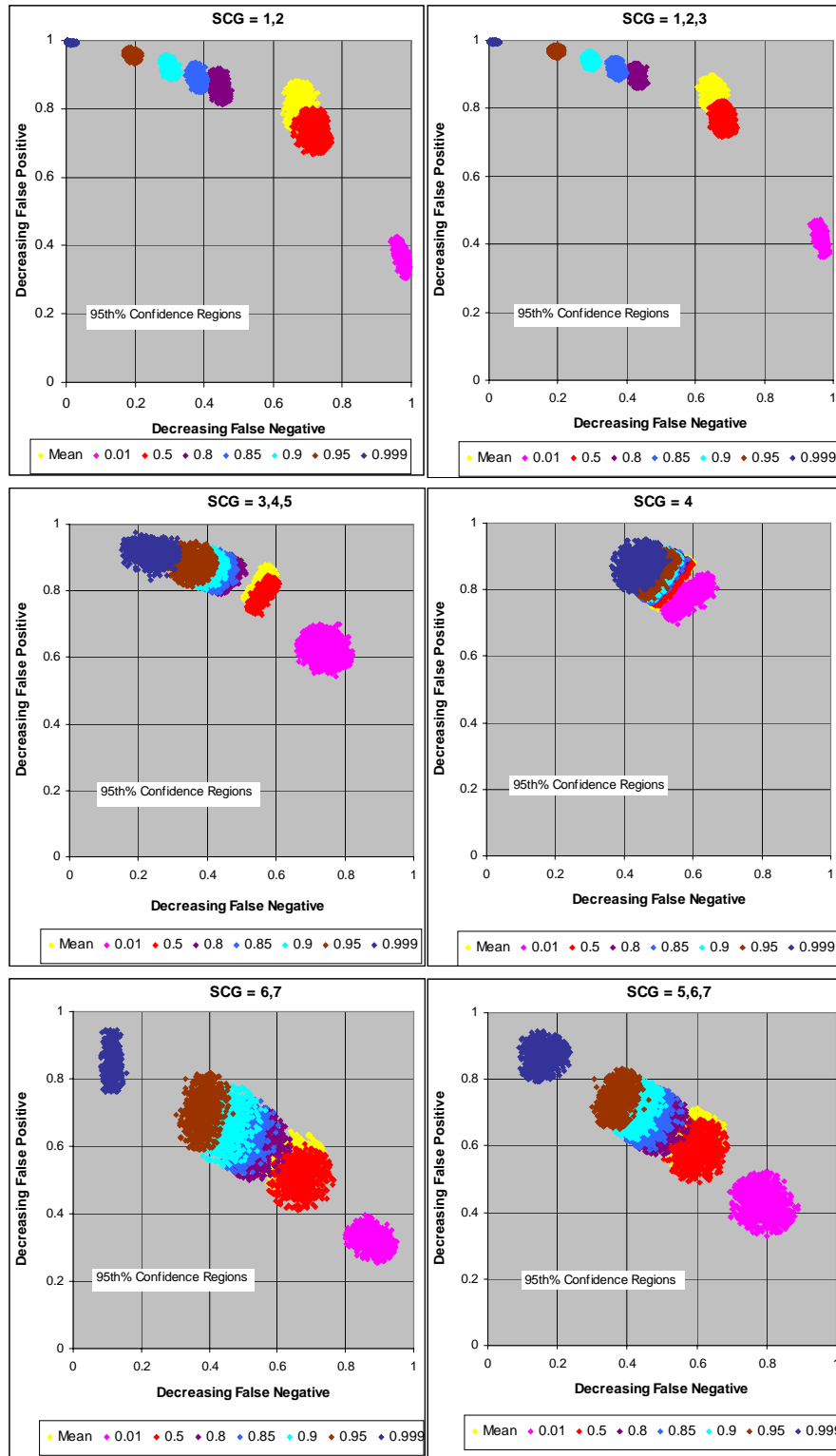


Figure E-14. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Summed Dosages and Based on Delaunay Triangulation Interpolation of Field Trials (Based on Summed Dosage on a Linear Scale)

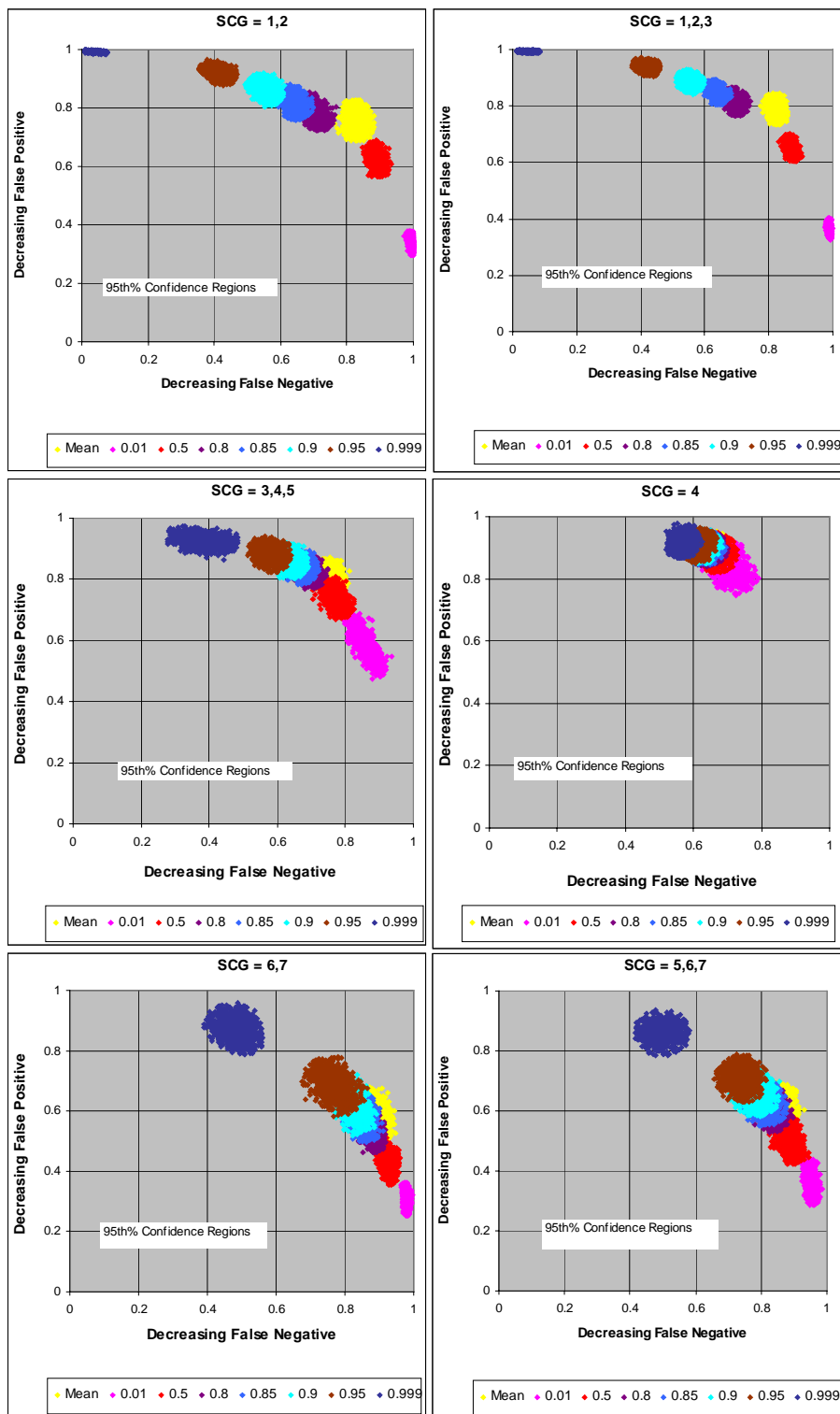


Figure E-15. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Summed Dosages and Based on Delaunay Triangulation Interpolation of Field Trials (Based on Summed Dosage on a Logarithmic Scale)

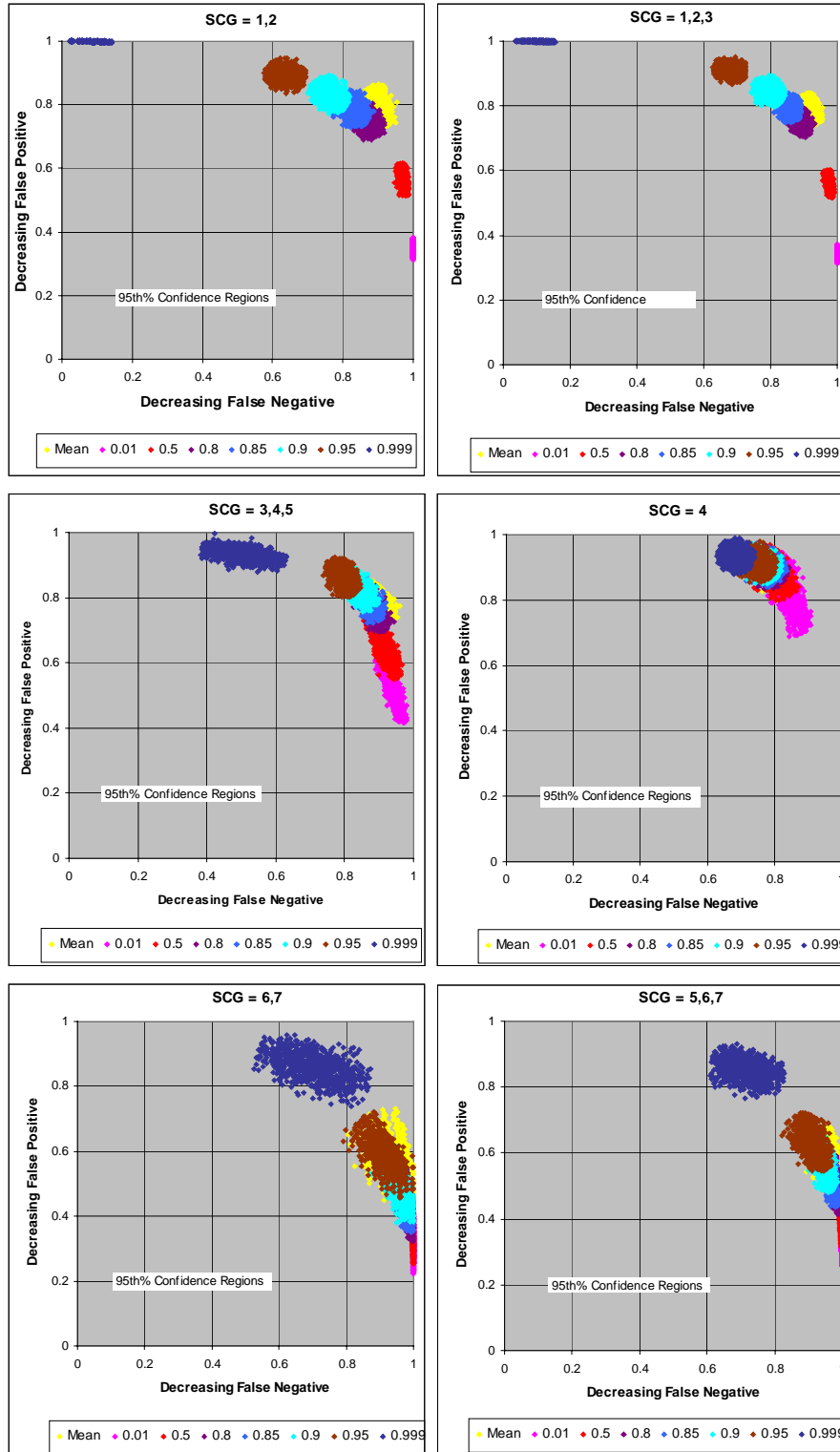


Figure E-16. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Threshold of 60 mg-sec/m³ and Based on Polar Bilinear Interpolation of Field Trials

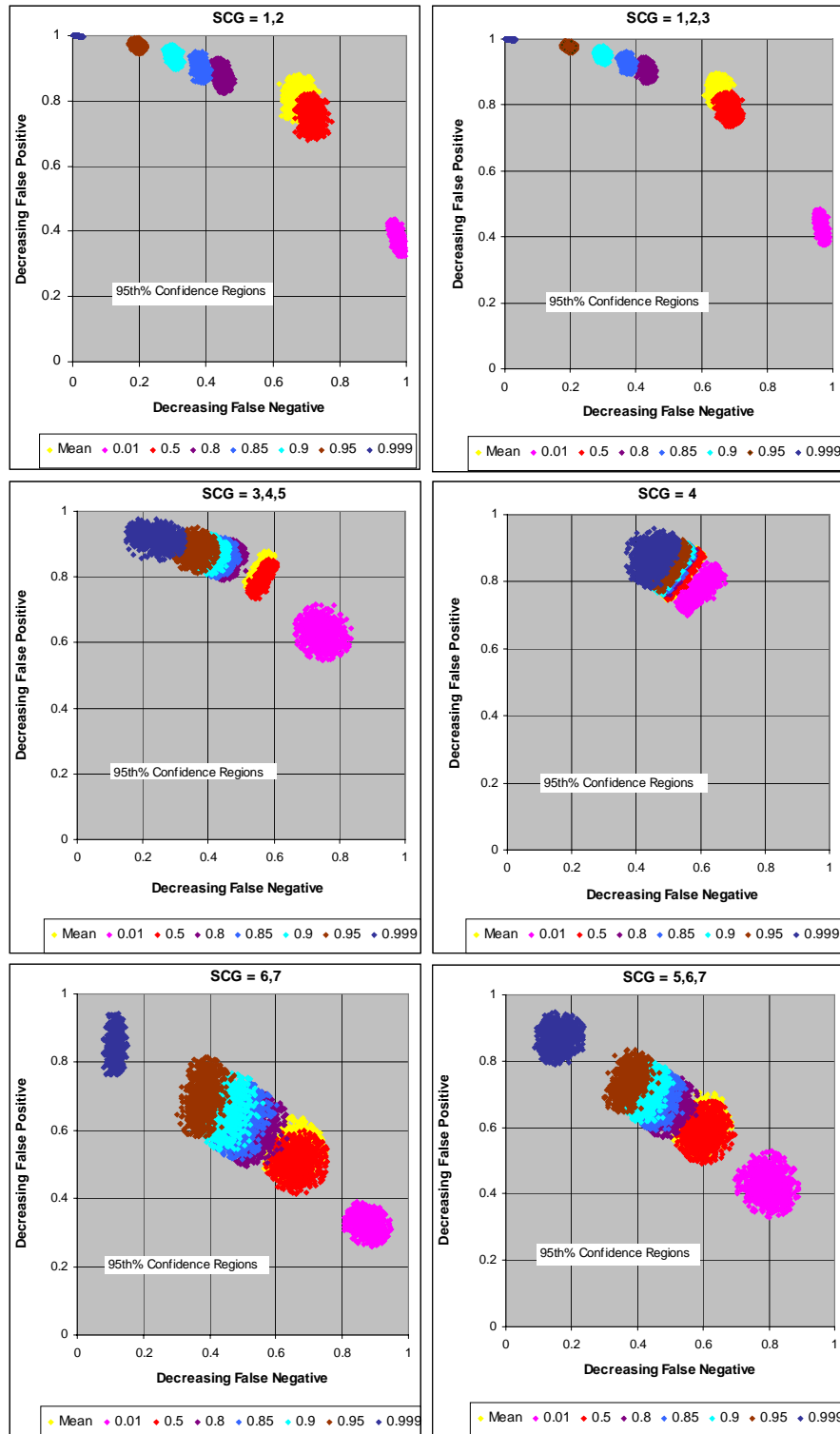


Figure E-17. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Summed Dosages and Based on Polar Bilinear Interpolation of Field Trials (Based on Summed Dosage on a Linear Scale)

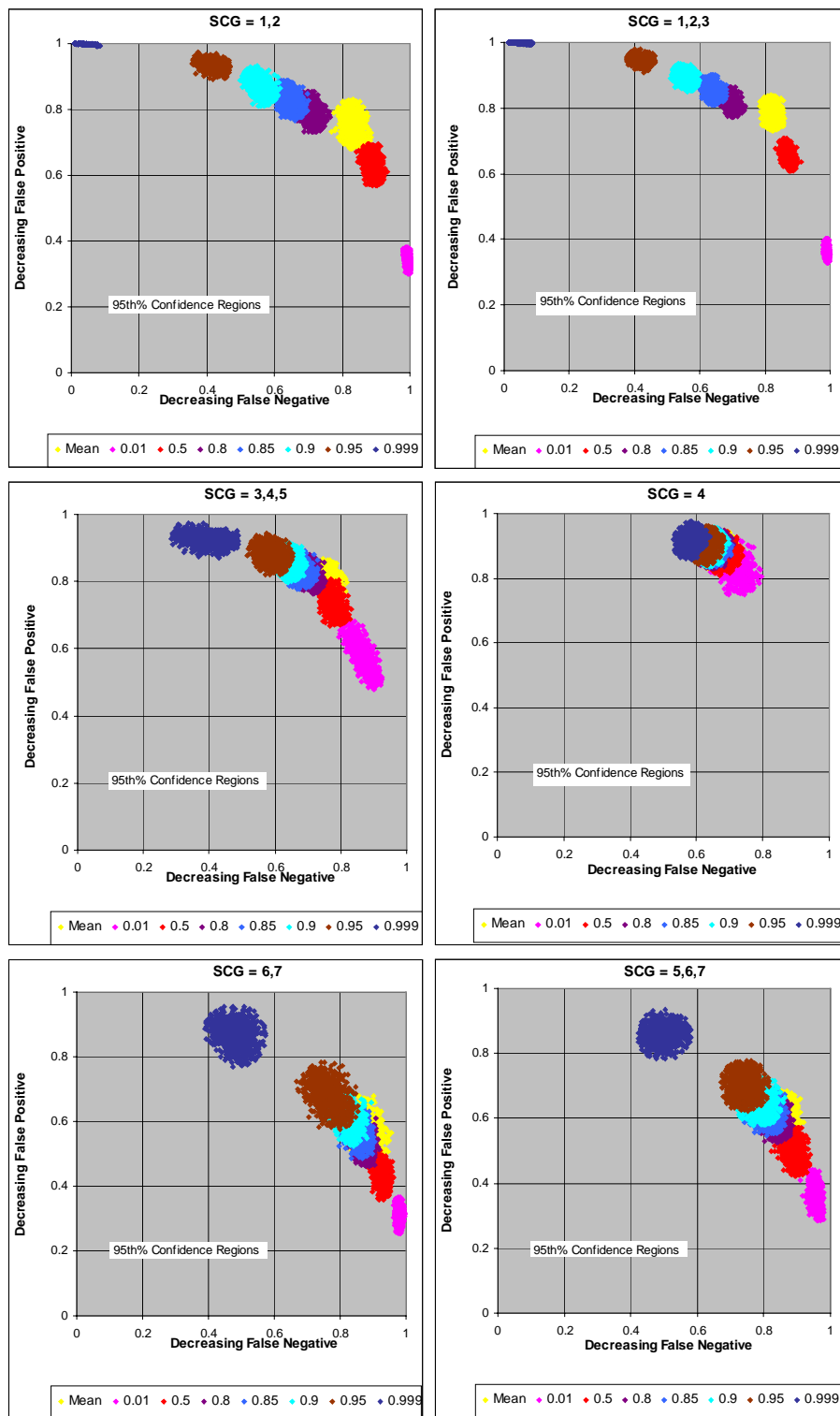


Figure E-18. MOE Confidence Regions for HPAC Conditional Probabilistic Prediction Outputs (and Mean Value): By SCG for Summed Dosages and Based on Polar Bilinear Interpolation of Field Trials (Based on Summed Dosage on a Logarithmic Scale)

REFERENCES

- E-1. Barad, M. L. (Editor), *Project Prairie Grass, A Field Program in Diffusion*, Geophysical Research Papers No. 59, Volumes I and II, DTIC #AD-152572/AFCRC-TR-58-235(I), Air Geophysical Laboratory, Hanscom Air Force Base, MA, 1958.
- E-2. Irwin, J. S. and Rosu, M-R., "Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models," *Proceedings of the 10th Joint Conference on the Applications of Air Pollution Meteorology*, American Meteorological Society, Boston, pp. 6-10, 1998.

APPENDIX F
TASK ORDER EXTRACT

APPENDIX F

TASK ORDER EXTRACT

DC-9-1797

TITLE: Support for DTRA and LLNL in the Validation Analysis of Hazardous Material Transport and Dispersion Prediction Models

This task order is for work to be performed by the Institute for Defense Analyses (IDA) under Contract DASW01-98-C-0067, for the Defense Threat Reduction Agency (DTRA) in the Department of Defense. This task order authorizes funding for FY 2000.

1. BACKGROUND:

The Hazard Prediction and Assessment Capability (HPAC) is a suite of codes that predicts the effects of hazardous material releases into the atmosphere and their impact on civilian and military populations. The software can use integrated source terms, high-resolution weather forecasts, and particulate transport models to predict hazard areas produced by battlefield or terrorist use of weapons of mass destruction (WMD), by conventional counterforce attacks against WMD facilities, or by military and industrial accidents. HPAC is a forward deployable, counterproliferation and counterforce capability software tool available for government, government-related, or academic use. This tool assists warfighters in selecting weapon mixes for targets containing WMD and in emergency response to hazardous agent release. HPAC's relatively fast-running, physics-based algorithms enable users to model and predict hazard areas and human collateral effects in minutes.

The DTRA Verification and Validation (V&V) Program represents ongoing activities performed in parallel with development of all predictive codes in support of HPAC. One element of V&V is to perform code-on-code comparisons. In this strategy, each code receives the same input. In this manner, differences in the output predictions can lead to the identification of software bugs, or help to assess technical strengths and weaknesses of component algorithms within each code. In addition, a certain amount of credibility for both models is achieved when their predictions agree. When the inputs are simple, such as for fixed winds and simple terrain, the predictions tend to be dominated by the dispersion algorithms. Comparisons at this level of complexity are important to establish fundamental dispersion algorithm veracity, and to help discover software bugs. As more complex terrain and weather is included as input, the number of physical processes

responsible for transport and dispersion increases and the predictions become the result of many interdependent algorithm calculations.

Code-on-code comparisons will be performed using the DTRA code HPAC, the Lawrence Livermore National Laboratory (LLNL) code Atmospheric Release Advisory Capability (ARAC), and, possibly, Sandia-developed codes. These codes represent major national investments in transport and dispersion modeling within their respective applications. The comparisons will provide information from which to validate the HPAC and ARAC models (and perhaps others), as well as provide an opportunity to advance both technologies. The code comparisons will include short, medium, and long-range transport distances. Complex terrain and weather will also be included.

It is very difficult to separate meteorological uncertainty from the transport and dispersion model accuracy when comparing predictions to field-trial validation quality or real-world data. The validation challenge is to assess whether a model performs well over different field trials, and ultimately reflects real-world phenomena. Some codes perform better under certain conditions and specific scenarios. Hazard prediction models are generally developed for a range of user communities and applications. Each user community has a different set of requirements. Thus, the corresponding hazard models tend to be optimized for specific applications. The process of accrediting a model is always couched in terms of the end-user requirements.

Various figures-of-merit (FOM) are used to express model performance relative to observed data. Most FOMs tend to use manifestations of a ratio (geometric or arithmetic) between the predicted and measured quantities. The compared quantities are usually peak, plume-centerline, and off-axis concentration or dosage, as well as crosswind and along-wind spread and area coverage. Other FOMs may include the second-moment of the dosage and concentration values at a sampler location. All these FOMs are reasonable validation performance measures, but none of them explicitly expresses an application-oriented performance measure. A “yardstick” is needed that measures application-oriented model performance. The scale on this yardstick would clearly and directly relate to specific user’s concerns and needs. The pursuit of this “accreditation” performance measure is a new initiative at DTRA.

2. OBJECTIVE:

IDA will conduct independent analysis and special studies associated with verification and validation of the suite of models associated with the Hazard Assessment and Prediction Capability. IDA will support development of user-oriented performance measures of effectiveness (MOE) using validation quality field trial data sets; coordinate scenario definition and arbitration for code-on-code V&V activities; and assist DTRA and the Department of Energy in identifying the V&V parameter space associated with various hazard assessment and collateral effects communities.

The objectives of verification and validation analysis and coordination are: (1) to ensure that a consistent analysis approach is used when comparing model predictions, and assist DTRA in the implementation of code-on-code analysis, comparisons, and interpretation; and (2) to define measures of effectiveness in terms of user-specific objectives and applications.

The scope of this effort may be expanded to other programs as directed by DTRA.

3. STATEMENT OF WORK:

As required by DTRA technical representatives, IDA will perform the following tasks:

a. Support the planning, implementation, arbitration, and evaluation of code-on-code comparison activities. The purpose of these activities is to compare the transport and dispersion algorithms and corresponding output predictions between DTRA's suite of models, the DOE ARAC's suite of transport and dispersion models, and other models as called for. IDA will support code-on-code scenario definition, coordinate the identification and implementation of common performance measures, and support development of a common analysis approach. IDA will conduct independent analysis, as needed, to support the code-on-code analysis and interpretation of the results.

b. Explore validation and accreditation MOEs given a framework that includes quantification of false positive and false negative predictions. This exploration would include the computation of MOE values for various formulations based on short-range comparisons of HPAC and ARAC predictions to field trial data. A key to interpreting the results of this effort will be obtaining a sense for what are the acceptable user requirements. These requirements will differ among potential user groups (military targeting, passive CB defense, civilian first responders, military versus civilian population human effects, etc.).

4. CORE STATEMENT:

This research is consistent with IDA's mission in that it will support specific analytical requirements of the sponsor and will assist the sponsor with planning efforts. Accomplishment of this task order requires an organization with experience in operationally oriented issues from a joint and combined perspective, which IDA, a Federally Funded Research and Development Center, is able to provide. It draws upon IDA's core competencies in Systems Evaluations and Operational Test and Evaluation. Performance of this task order will benefit from and contribute to the long-term continuity of IDA's research program.

REPORT DOCUMENTATION PAGE*Form Approved*
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22203-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)**2. REPORT DATE**

May 2001

3. REPORT TYPE AND DATES COVERED

Final: Dec. 2000 – May 2001

4. TITLE AND SUBTITLEApplication of User-Oriented MOE to HPAC Probabilistic Predictions of *Prairie Grass* Field Trials**5. FUNDING NUMBERS**DASW01-98-C-0067, OUSD(A)
DC-9-1797**6. AUTHOR(s)**

Dr. Steve Warner, Dr. James F. Heagy, Dr. Nathan Platt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311**8. PERFORMING ORGANIZATION REPORT NUMBER**

IDA Paper P-3586

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)Dr. Allan Reiter
Defense Threat Reduction Agency (DTRA)
6801 Telegraph Road
Alexandria, VA 22310-3398**10. SPONSORING/MONITORING AGENCY REPORT NUMBER****11. SUPPLEMENTARY NOTES****12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution unlimited. Directorate for Freedom of Information and Security Review, 22 May 2001.

12b. DISTRIBUTION CODE**13. ABSTRACT (Maximum 200 words)**

The goal of this task is to improve the verification, validation, and accreditation (VV&A) of hazard prediction and assessment models. These studies are part of a larger joint VV&A effort that DTRA and the Department of Energy, via the Lawrence Livermore National Laboratory (LLNL), are conducting. The Institute for Defense Analyses' role includes conducting comparisons between models, providing analysis and discussions associated with these examinations, and exploring and developing measures of effectiveness (MOE) that can aid hazard prediction model validation and accreditation. This paper develops and demonstrates concepts that can ultimately aid validation and user accreditation of transport and dispersion models. The paper applies a user-oriented MOE to the probabilistic outputs that are obtained from the Hazard Prediction and Assessment Capability (HPAC) model predictions of the short-range *Prairie Grass* field trials. This MOE allows for the assessment of the probabilistic predictions in terms of the false negative and false positive fractions. Quantitative descriptions of a user's potential risk tolerance are also described in this paper. This study develops these quantitative descriptions and uses them to evaluate the "goodness" of HPAC predictions of the *Prairie Grass* field trials from the perspective of a few notional users.

14. SUBJECT TERMSModel validation; hazardous material transport and dispersion; HPAC; probabilistic outputs; measure of effectiveness; *Prairie Grass***15. NUMBER OF PAGES**

300

16. PRICE CODE**17. SECURITY CLASSIFICATION OF REPORT**

Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT

Unclassified

20. LIMITATION OF ABSTRACT

Unlimited